The Dissertation Committee for Clare Butler Schoene
certifies that this is the approved version of the following dissertation:

**Electrical Parameter Control for Semiconductor Manufacturing**

Committee:

S. Joe Qin, Supervisor

Thomas Edgar, Supervisor

Erhan Kutanoglu

John Hasenbein

Grant Willson

Gyeong Hwang

John Stuber

# Electrical Parameter Control for Semiconductor Manufacturing

**by**

**Clare Butler Schoene, B.S.**

**DISSERTATION**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2007

# Acknowledgments

First I would like to thank my advisor, Dr. Qin, for his guidance, help, and advice through this research project. I also want to thank the Qin and Edgar research groups for their friendship and support and for keeping life on the fifth floor interesting. I also would like to thank John Stuber for helping me and being my mentor and friend while I was at Texas Instruments as well as at school. He helped make my research relevant and useful to industry and provided me with real manufacturing data to support my research efforts.

I also want to thank all my friends and family that have helped me through this whole experience. Libby for keeping me going through the up and downs, Dad for always believing I could do it, Katura for always being there, and Adam for giving me confidence in myself and my abilities and helping me get through the very end.

# Electrical Parameter Control for Semiconductor Manufacturing

Publication No. ______________

Clare Butler Schoene, Ph.D.
The University of Texas at Austin, 2007

Supervisors: S. Joe Qin
Thomas Edgar

The semiconductor industry is highly competitive environment where modest improvements in the manufacturing process can translate to significant cost savings. An area where improvements can be realized is reducing the number of wafers that fail to meet their electrical specifications. Wafers that fail to meet electrical specifications are scrapped, which negatively impacts yield and increases manufacturing costs. Most of the existing semiconductor process control research has focused on controlling individual steps during the manufacturing process via run-to-run control, but almost no work has looked at directly controlling device electrical characteristics. Since meeting electrical specifications is so critical to reducing scrap a fab-wide electrical parameter control scheme is proposed to directly control electrical parameter values. The goal of the controller is reducing the variation in the electrical parameters. The control algorithm uses a model to predict electrical parameter values after each processing step. Based on this prediction the decision to make a control move is made. If a control move is necessary, optimal adjustments for the subsequent processing steps are determined. The process model is continually updated so that it reflects the current process. A simple implementation using a least squares model is first proposed. Simulations and an industrial case study demonstrate the potential improvements that can be achieved with the algorithm and the limitations of the simple implementation are discussed. A partial least squares modeling and control algorithm combined with missing data algorithms are proposed as enhancements to the electrical parameter control algorithm to address many of the issues faced when implementing such a control strategy in real manufacturing environments. The enhancements take the input variable correlations into account when making control moves and utilize the correlation structure to make better model predictions. Simulations are performed to determine the effectiveness of the enhancements. A cost function formulation and a Bayesian based alternative are also presented and evaluated. The cost function implementation uses a different method to determine the optimal set points for the subsequent processing steps than the other implementations use. Simulations are used to compare the cost function formulation with the other methods presented. The Bayesian implementation addresses the stochastic nature of the manufacturing process by dealing with the probabilities of events occurring. A simulation of the Bayesian algorithm is preformed and the algorithms limitations are discussed.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

The semiconductor manufacturing industry is constantly faced with the challenge of producing ever-smaller devices while continuing to remain profitable. This goal has necessitated the move from 200mm to 300mm diameter wafers where manufacturing efficiency can be improved and manufacturing cost per unit area of silicon can be reduced. In the mean time, the critical dimensions of VLSI devices keep shrinking, making it a constant challenge to improve yields and maintain high throughput. In order to justify the capital expenditures for 300mm fabrication facilities (fabs), maintaining operations efficiency and high yields is critical. The International Technology Road Map for Semiconductors has identified factory information and control systems as critical technology for attaining high yield and cost effectiveness in fabrication facilities [36]. The increase in weight of 300mm wafers requires the 300mm fabs to have automated systems for transporting wafers between equipment in the fab. The automated transportation requirements in 300mm fabs will facilitate the use of other automated technologies creating more potential for the integration of control and information systems as well.

Major factors affecting yield, cycle time, and manufacturing costs are the machine time lost and material wasted when wafers need to be scrapped. Wafers can be scrapped for many reasons at various points in the manufacturing process and one of these critical decision points is during wafer electrical testing. Electrical testing is performed at the end of processing and electrical data is taken for each wafer. The electrical test data taken at parametric test are compared to product specifications to determine if wafers should be scrapped. Large variations in electrical parameter measurements translate to more wafers needing to be scrapped. Reducing electrical parameter variations creates cost savings because scrapping fewer wafers saves on materials costs and increases equipment utilization time. Since electrical parameter values directly affect yield, effectively controlling the electrical parameter variation is an important aspect of semiconductor manufacturing.

The general methodology for controlling electrical parameters has been aimed at controlling the individual manufacturing processes that influence the electrical characteristics of the devices being produced. The idea being that reducing the output variations of each manufacturing step will decrease the variations in the electrical parameters. This concept has helped to maintain electrical parameter variations at manageable levels and very little research has looked at other methods for improvement. The focus of this research is on the design of a controller that reduces electrical parameter variations by controlling the electrical parameters directly. This will result in increased efficiency and cost savings in manufacturing.

## 1.1 Device Electrical Characteristics

A 300mm wafer that has completed the manufacturing process contains billions of transistors [35]. A diagram of a metal-oxide-silicon (MOS) transistor is pictured in Figure 1.1. The source and the drain voltages applied to their respective electrodes, in combination with the gate voltage, are used to turn the transistor on and off by regulating the flow of current across the gate channel, represented in the diagram by $L$. The thickness of the oxide, $d$, plays an important role in device functionality because it directly affects the ability of the oxide to act as a capacitor. The two regions labeled $n^+$ are the doped regions that affect the charge balance in the transistor. Two major electrical characteristics of MOS devices are the threshold voltage and the drain current. Threshold voltage is the minimum voltage that needs to be applied to the gate for current to flow in the transistor and drive current is the current that flows through the channel between source and drain when the transistor is on. It is easy to see how the physical dimensions and of the device and the doping concentrations affect the electrical properties of the transistor when looking at the characteristic equations. The characteristic equation for threshold voltage, $V_T$, is shown in Equation 1.1

$$\begin{aligned} V_T &= \Phi_{ms} - \frac{Q_i}{C_{ox}} - \frac{Q_d}{C_{ox}} + 2\phi_F \qquad (1.1) \\ C_{ox} &= \frac{\epsilon}{d} \end{aligned}$$

and the equation for drive current, $I_D$, is shown in Equation 1.2 [87]

$$I_D = \frac{\mu_n Z C_{ox}}{L}[(V_G - V_T)V_D - \frac{1}{2}V_D^2]. \tag{1.2}$$

In the equation for threshold voltage $\Phi_{ms}$ is the metal semiconductor work function potential difference, $Q_i$ is the effective MOS interface charge, $C_{ox}$ is the oxide capacitance, $Q_d$ is the depletion region charge, $\phi_F$ is the difference between the intrinsic and the Fermi energy levels, $\epsilon$ is the permittivity of the oxide, and $d$ is the oxide thickness. In the equation for drain current $\mu_n$ is the electron mobility, $Z$ is the width of the channel in the $z$ direction, $L$ is the channel length, $V_G$ is the gate voltage, and $V_D$ is the drain voltage. Even in the most basic MOS characteristic equations it is clear that the physical dimensions of a device, such as oxide thickness and gate length, have a direct impact on the device's electrical performance. The doping concentrations of the substrate, as well as the channel, affect the depletion region charge and also have significant impact on electrical performance. As the complexity of manufactured devices increases and device size decreases other geometries and processing parameters begin to affect electrical performance as well. Devices manufactured today also have many other electrical parameters of interest. These include leakage current, which is especially important for mobile applications, ring oscillation frequency, which is a multi-transistor structure that tests gate delay time for the whole wafer and is not device specific, parasitic capacitances and resistances, and subthreshold slope.

The device geometries that are important to electrical functionality are patterned onto the wafer during manufacturing. Since it difficult to measure these geometries for every transistor test structures designed for measurement are also patterned onto the wafer. The test structures are measured after the manufacturing process that defines them. Some of these measurements are known as critical dimensions (CDs) and others are just known as metrology data. A detailed description of the manufacturing process and all of the major discrete processing steps that comprise it are outlined below, along with the CD and metrology measurements and measurement methods that are associated with them. Further details can be found in [38, 62, 65].

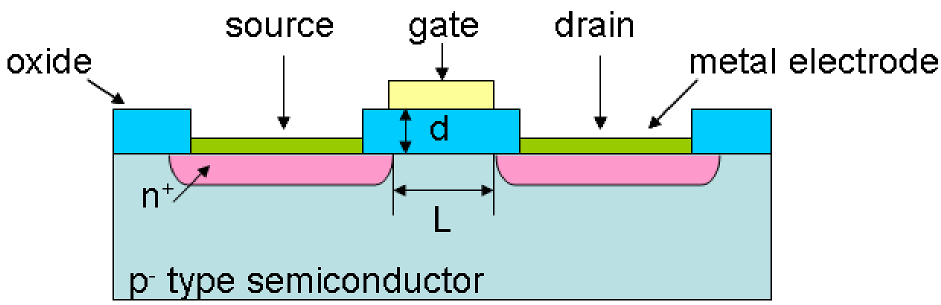


Figure 1.1: Diagram of a basic MOS transistor [89]

## 1.2 Semiconductor Manufacturing Process

The semiconductor manufacturing process for logic devices is a complex process consisting of hundreds of discrete processing steps. These steps are performed in series, culminating in electrical tests, before the chips are cut and packaged for use. Blank wafers are passed through the major steps of deposition, lithography, etch, doping, and chemical-mechanical planarization

(CMP) processes multiple times as layers of interconnects are patterned onto the surface of the wafers. The deposition step is responsible for film formation on the wafer surface followed by lithography. Lithography is followed by etch and the combination of the two steps results in the circuit pattern transfer to the surface of the wafer. Once the pattern is transferred the wafer is doped in selective regions and another film is put on. CMP is used to level the surface. This may need to be done before other transistor features are patterned on, before metal layers are deposited, or before layers of interconnects are formed. Figure 1.2 shows the basic manufacturing steps in semiconductor fabrication and Figure 1.3 shows a series of diagrams depicting how the transistor is patterned onto the wafer surface by undergoing the different processes involved in manufacturing. Despite the simplicity of the manufacturing flow shown in Figure 1.2 the actual wafer flow through a fab is complex. Each wafer follows a different path because there are multiple tools that perform the same process and each wafer may go through certain processing steps more than once. This is because the same process may need to be performed at different points in the manufacturing flow. Also adding to the complexity of the process is the fact that most of the major processing steps consist of many steps themselves. An example is lithography, that involves photo resist application, exposure, development, etch, and photo resist removal steps. Combining all of these steps over multiple layers results in upward of 300 processing steps to manufacture most of the logic devices that are in use today.

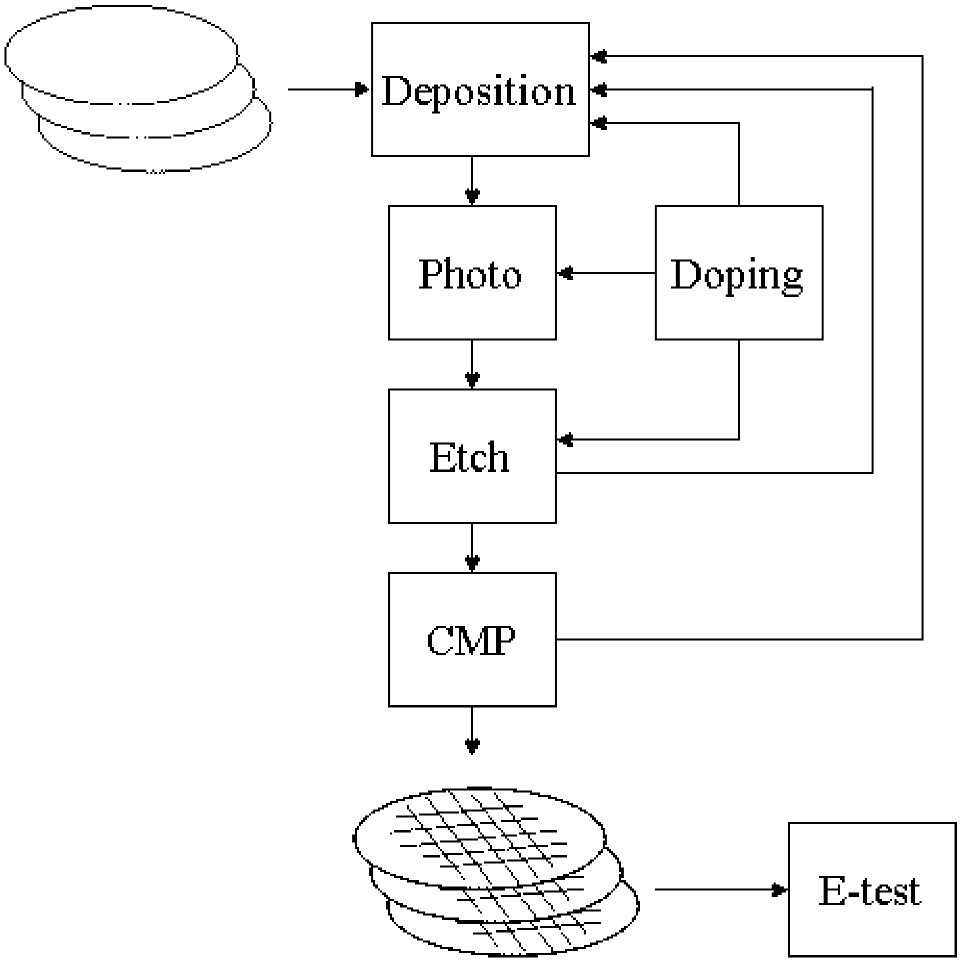


Figure 1.2: Basic steps in a semiconductor manufacturing process [2, 52]

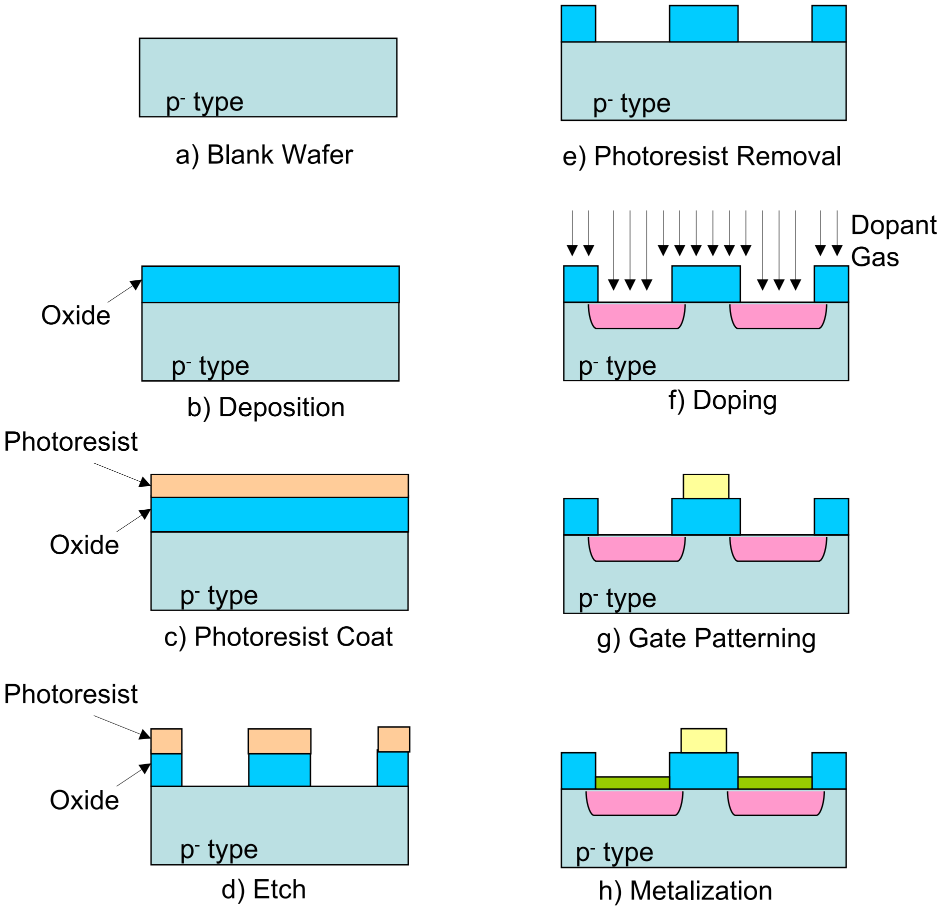


Figure 1.3: MOS fabrication sequence a) Blank p-type silicon wafer b) Deposition is used to put an oxide layer on top of the silicon c) The lithographic sequence starts by coating the wafer in photoresist d) The photoresist is exposed and the unmasked regions are etched away e) The remaining photoresist is removed f) The channel pockets are formed by doping g) The gate is patterned onto the oxide h) The metalization step forms the source and drain contacts

### 1.2.1 Deposition

Deposition is the process of depositing material on the surface of a wafer. This is accomplished by growing thin films through processes like chemical vapor deposition and physical vapor deposition, or coating the wafer, as in molecular beam epitaxy and electrochemical deposition. Deposition is used to deposit thermal oxides, dielectric layers, polycrystalline silicone, and metal films, all of them playing critical roles in device production and functionality. Most deposition tools process wafers in groups, or lots, and adjustment of the process for individual wafers is not possible. The critical dimension measurement associated with deposition is the thickness of the deposited material. Ellipsometry is a technique used to measure film thickness. The thickness measurement is based off of the changes in polarization of light reflected off the surface of the material. In most manufacturing processes either a few wafers are measured to determine the thickness CD or a separate test wafer is measured to check the CD value and monitor the process.

### 1.2.2 Lithography

Lithography is the process of patterning the wafer with the geometric features that form transistors, such as implant regions and contact windows, and as mentioned earlier, involves many steps. A mask of the pattern to be transferred is used to expose the uncovered portion of a wafer coated in photoresist. The chemical properties of photoresist change during exposure so the pattern specified by the mask is transferred to the wafer by flooding the wafer in a developer solution that removes the exposed photoresist, revealing the insulation layer below. An etch process that etches the insulation layer, but not the remaining photoresist, is used to transfer the pattern into the insulation layer. Once the etching is complete the remaining photoresist is stripped away leaving an insulation layer patterned like the mask. If a negative photoresist is used then the pattern that is transferred is the inverse of the mask. Depending on the type of equipment used wafers can be processed individually or in lots. Photoresist is usually spin coated onto the wafer and this is done on a wafer to wafer basis. Baking and etch steps can be performed on individual wafers or lots and the exposure step is always performed on each wafer individually. After the lithography process is complete the size or dimensions of the features patterned onto the wafer are measured. One of the most important CDs determined during lithography, in terms of electrical functionality, is the gate length pictured in Figure 1.1. This feature is difficult to measure and the gate conductor width of a test strucure is often measured instead. This measurement is made using a scanning electron microscope (SEM), which uses electron wave lengths to create detailed images of the patterned features. Measurements of other repeated line widths patterned during lithography are made using scatterometry, which uses the pattern of light diffracted from the patterned surface to determine the size of the features. Usually scatterometry measurements are used whenever possible because the SEM measurements take a long time and require the wafer to be cut so that it is destroyed in the process. Scatterometry tools measure each wafer individually but usually only a few of the wafers in each lot are measured to reduce testing time.

### 1.2.3 Etch

Etching is the process of removing part, or all, of a chemical layer on the surface of a wafer and can be preformed in two ways, wet etching and dry etching. Wet etch is useful for blanket etches of the entire wafer. Dry etching is more suitable for removing patterned material and is the type of etching used after lithography. Plasma etching, reactive ion etching, and sputter etching are all dry etching methods. The etching process is performed on individual wafers. The metrology measurement taken upon completion is the depth of the etch. For the blanket etch case this is the thickness of material removed during the etching process. For patterned etches the etch depth is measured and sometimes the sidewall angle is measured as well. The sidewall angle measures how close to perpendicular the wall of the groove formed during etching is to the surface of the wafer. The metrology measurements measurements are usually only made for a few of the wafers in each lot.

### 1.2.4 Doping

Doping is the process of adding known quantities of ion impurities to wafers to change their electrical characteristics. This process is performed many times in different regions of the wafer throughout manufacturing. Doping is performed through ion implantation and diffusion. Ion implantation uses an ion beam to implant the impurities and the doping distribution is mainly determined by the mass and energy of the implanted ions. Diffusion processes expose the wafer surface to the dopant at high temperatures and the dopant distribution is primarily a function of the temperature and diffusion time. The doping process is performed on one wafer at a time and there are no concentration measurements that can be made on product wafers. The sheet resistance on pilot wafers can be used to approximate the doping concentration but this is usually only done for calibration purposes.

### 1.2.5 Chemical-Mechanical Planarization

Chemical-mechanical planarization is a process used to create a flat surface across the entire wafer. The planarization of the wafer surface allows multiple levels interconnects to be patterned onto the wafer and is also effective at reducing defect density. The CMP process moves the wafer against a pad, with a slurry between the pad and the wafer surface that removes material, leveling the surface. CMP utilizes mechanical material removal via abrasive particles in the slurry and chemical material removal by agents in the slurry. The result is a much more uniform and defect free surface than is achievable through mechanical methods alone. CMP tools can process anywhere from one to five wafers at a time. The metrology measurements that characterize the CMP process are the thickness of the planarized surface and the surface uniformity.

## 1.3 Manufacturing Data and Availability

The types of data collected during manufacturing, and the availability of the data for use, are important factors in designing control strategies for semiconductor manufacturing processes. A wide array of different data types, including wafer state data and equipment state data, are collected during manufacturing. Wafer state data consists of measurements related to the physical parameters of the wafer for each processing step and include the CD and metrology measurements discussed in the previous sections. Equipment state data are data collected on the processing parameters of the equipment, such as temperatures, pressures, flow rates, and processing times. During most processing steps equipment state data is collected on the processing variables in order to monitor, and sometimes control, the process. After the processing step is complete metrology data are measured to determine the state of the wafer, including the CD values, for the completed step.

Due to the large number of processing steps, the cost of testing, and the time it takes to make metrology measurements, metrology data for each wafer may not be collected during manufacturing. Instead, only a subset of wafers from each lot may be measured or a non-production wafer, also known as a pilot wafer, may be included with the lot. In the latter situation, the pilot wafer is measured instead of the wafers in the lot and that value is used to monitor the process. In some situations metrology data may be delayed due to the nature of the measurements being performed or no metrology data may be available at all, as is the case with ion implantation. The lack of continuity in metrology measurements for a particular wafer from each processing step creates many challenges during model building as well as control algorithm development. These issues are addressed in this document so that the resulting control strategy is applicable to real manufacturing processes.

## 1.4 Current Control Strategies

There are multiple layers of control that occur during semiconductor manufacturing. The first layer of control is real time control at the equipment level. Each of the processing steps described in the previous section are complex chemical reactions. The equipment that performs these processes have real time controllers that adjust temperatures, pressures, flow rates, and other processing variables to ensure that the chemical reactions proceed as planned in a safe, and somewhat consistent, manner. The equipment manufacturers develop and tune the controllers and they are embedded in the tools such that no work is needed on the part of the fab operator to implement or tune them before the process can be successfully run on the tool. This level of control is inherent in the system and operates independently for each wafer or batch of wafers processed by the tool.

The next layer of control is implemented to achieve more consistency in the process output, or metrology measurement, of each of the processing steps by utilizing the fact that each wafer, or batch of wafers, is not independent. This was first accomplished through statistical process control (SPC). SPC is not actually a controller, but a method for monitoring a process so that adjustments can be made when the process is deemed "out of control." Research over the last 15 years has led to the development of run-to-run (R2R) control, which takes a more active role by making process adjustments before the process has reached an "out of control" state. R2R controllers are discussed in detail in Section 2.1. Most fabs in operation today have both SPC and R2R controllers in operation, depending on the requirements of the processing step that is being controlled. Either way, the goal of the second layer of control is to maintain the CD, or other geometrical features, at their targeted values. This type of control is process specific and is developed and implemented by engineers in the fab. This layer of control uses information from the previously processed wafers, or batches of wafers, when making process adjustments but is independent of the upstream or downstream processes in the manufacturing flow.

The goal of the third layer of control is to achieve CDs and other geometric features that are closer to target by accounting for the upstream and downstream processes when making adjustments to the current process. In this layer information from a completed process is fed forward to the next process so that control adjustments at the next step can use the information from the previous step, as well as the information about previously processed wafers at the next step, when making process adjustments. This layer is also developed and implemented by engineers in the fab and is referred to as module level control.

The focus of the different layers of controllers described so far is the same: keeping geometric features at their targeted values. This methodology indirectly works to keep electrical parameters on target because the geometric features are consistently sized, but there is room for improvement. The current controllers may be optimal at every step in the manufacturing process but the optimality of each individual step does not always yield the optimal solution for the entire process. Poorly tuned controllers, switching between products, and metrology errors all contribute to manufacturing errors at each processing step. As a result, small errors in each processing step can accumulate in up to 300 processing steps to produce large errors and even off-specification products, in terms of electrical parameters. To avoid the accumulation of errors across many manufacturing steps coordination and the sharing of information among these sequential steps must be considered. Fab-wide control strategies have been proposed to address and facilitate coordination for the entire manufacturing process [55, 67, 71]. In fab-wide control the focus is shifted from controlling the individual manufacturing steps to controlling end of processing parameters of interest, such as yield or electrical parameters. Fab-wide controllers can make adjustments to later processing steps to account for errors introduced in earlier processing steps in order to avoid severely off target electrical parameters. Fab-wide controllers also have the added benefit of compensating for metrology drift not handled by R2R control [14]. Work done by Harrison et al. [34] showed that a fab-wide controller could significantly reduce the variation in the erase time for a flash memory device. Fab-wide control has shown great potential but no detailed and implementable control scheme has been developed to date. The goal of this dissertation is the development of a fab-wide control algorithm that coordinates lower level R2R controllers in order to reduce variability and improve overall quality in terms of electrical parameters. The algorithm has enough detail and flexibility that it may be implemented in a manufacturing environment and issues relating to implementation are discussed. This algorithm has the potential to increase yield and provide cost savings by reducing variability and scrap during semiconductor manufacturing.

## 1.5 Dissertation Outline

Now that the general goals of this dissertation have been discussed and the basic manufacturing environment that the algorithm must work within has been described the rest of the dissertaion is outlined as follows. Chapter 2 presents a detailed literature review of the relevant research in semiconductor manufacturing control, including R2R control, module level control, and fab-wide control. There is also a review of other areas of research that are applicable to the control strategies developed in this work. Chapter 3 discusses how the electrical parameter controller algorithm works and the three major components that make up its design. Chapter 4 develops a simple, but detailed, implementation of the EPC algorithm. Simulations and an industrial case study are used to demonstrate the potential for the EPC strategy and an assessment of the limitations of the particular implementation is made. Chapter 5 discusses enhancements to the EPC algorithm to address many of the limitations outlined in Chapter 4. Simulations are used to demonstrate the effectiveness of the enhancements. Chapter 6 presents two alternative implementation strategies. The first is a different objective function formulation for determining optimal set point calculations and the second is a Bayesian approach that addresses the stochastic nature of manufacturing processes by dealing with probabilities instead of deterministic values. Chapter 7 summarizes the work presented and offers directions for future work.

# Chapter 2

# Literature Review

The different types of control that are implemented in a semiconductor manufacturing fab are real-time equipment control, R2R control, and module level control. A general description of each of these controllers is outlined in Section 1.4. Real-time equipment controllers are tool specific controllers that are responsible for controlling the chemical processes occurring within a tool. They are embedded into the software and hardware that comes with the processing equipment and are not adjustable or configurable by the tool user. R2R controllers, module level controllers, and the control algorithms presented in this dissertation are developed under the assumption that each of the processes in the fab have real-time equipment controllers and they are functioning normally. Real-time equipment controllers are generally simple feedback control loops and they are not reviewed any further in this chapter. The development or R2R control, module level control, and other algorithms that are important to the work presented in this dissertation are discussed in the following sections.

## 2.1 R2R Control

Much of the research in semiconductor manufacturing control has focused on run-to-run (R2R) control and it is the most commonly used advanced control strategy in industry. R2R controllers act on an individual manufacturing step to better control the quality measurement, or CD value, for that step. This type of controller seeks to maintain output parameters at their specified values by adjusting the recipe for the manufacturing process. The adjustments are based on metrology data at the equipment level. Run-to-run control is an attractive control strategy because it can be successfully applied when there is no in-situ process information available during processing [5] and because of the ability to apply it to any tool regardless of the manufacturer. Examples of areas where R2R control has been applied can be found in [19, 33, 84, 100].

The first R2R controllers were an automated extension of the manual adjustments made from SPC techniques. Examples of this are works by Guldi [32] for controlling oxidation thickness and Leang and Spanos [46] for controlling the thickness and concentration of photoresist. Guldi's controller only makes an adjustment to the oxidation time when the oxidation thickness sets off an SPC alarm. Once there is an alarm the adjustment is calculated using a linear model relating the oxidation thickness and the oxidation time. Leang and Spanos use an SPC based multivariate cumulative sum for control alarm generation and use a response surface model to make adjustments. The most widely used R2R controller combines commonly used (SPC) techniques and advanced process control methodologies by monitoring the control moves and using a model to adjust processing parameters [48]. This method is proposed as the gradual mode controller by Sachs et al. [78]. Instead of making adjustments based off of control alarms Sachs et al. uses a controller that continually makes small adjustments to the process. The process model is also continually adjusted based off of previous data by weighting the historical data with an exponentially weighted moving average (EWMA).

The EWMA R2R controller maintains a process output at its specified value by adjusting the inputs to the manufacturing process. The inputs to the manufacturing process are the tool states, or processing variables, that are adjustable and directly related to the output measurements made at the end of processing. The processing parameter adjustments are made based off of the output metrology data from the previously processed wafer. EWMA R2R controllers rely on a model relating the processing inputs to the processing outputs to make the necessary process adjustments. In its simplest form the process model is a linear model relating a processing parameter, $u_t$, to an output measurement, $x_t$ [56]

$$x_t = Bu_t + a_t. \tag{2.1}$$

In this formulation, $B$, is the model gain and $a_t$ is the model bias term. The processing parameter adjustment that is needed to keep the output on target is determined by entering the target value for the output into the linear model and calculating the processing parameter value that yields the targeted output value. The calculation is shown in Equation 2.2

$$u_t = \frac{x_{t,target} - a_t}{B}. \tag{2.2}$$

The model is continually updated to reflect changes in the equipment from run to run. It is the model updating that allows EWMA R2R controllers to compensate for nonstationary processes. The model is updated after a wafer has completed processing by using the measured output value for the completed wafer and an exponentially weighted moving average to recalculate the model bias, as shown in Equation 2.3

$$a_t = \lambda(x_{t-1} - Bu_{t-1}) + (1 - \lambda_{R2R})a_{t-1}. \tag{2.3}$$

Here $a_{t-1}$ is the model bias term from the previous run and $\lambda_{R2R}$ is a weighting factor that ranges between zero and one. Small $\lambda_{R2R}$ values are more conservative and weight the previous model bias term more heavily than the new data and vice versa for large $\lambda_{R2R}$ values.

The use of a model to continuously make adjustments to the processing parameters has provided improvements not achievable with SPC alone [39]. This is because of how effectively the EWMA controller compensates for nonstationary processes as compared to SPC. Drifting processes are very common in semiconductor manufacturing due to the aging of equipment over time. EWMA controllers do not preform well in the presence of deterministic time dependent drift because the EWMA controller is not a minimum variance controller for a drifting process [12]. This has led to the development of the double EWMA controller, or predictor corrector controller (PCC), proposed by Butler and Stephani [5]. The double EWMA controller uses two EWMA formulas, one for the model bias and the other for estimating the drift rate. The PCC controller is modified by Chen and Guo [9] to adjust for the age of the process. This allows the double EWMA controller to be applied to situations where the data samples are not equally spaced in time. The use of and EWMA controller where the model was selected based on the machine, process, and material being used was proposed by Sullivan et al. [88]. EWMA based R2R controllers effectively handle many of the process disturbances common in real manufacturing environments which has led to their widespread use in industry. This, in turn, has led to the use of the term R2R control to be synonymous with EWMA based R2R control in many discussions.

The success of R2R controllers is based on the timely availability of metrology data. As online metrology capabilities were developed, R2R controllers made use of the newly available data for process control. For plasma etching Bushman and Farrer [4] use scatterometry measurements for poly-gate etch process monitoring and optical emission interferometry is used for end-point detection in polysilicon plasma etch processing by Wong et al. [99]. New CMP metrology is used for R2R control by Smith et al. [83] and temperature control is performed using diffusive reflectance spectroscopy by Wang et al. [94].

New algorithms and modeling techniques were incorporated into R2R control in order to address a wider array of the complex chemical processes present in semiconductor manufacturing. An adaptive R2R controller for plasma etch is proposed by Moyne et al. [58]. Fuzzy logic and a database learning mechanism are incorporated into the R2R controller so it can easily adapt to changing processes. Del Castillo and Yeh [15] develop an adaptive R2R controller for linear and nonlinear processes and Qin et al. [70] propose an adaptive R2R controller for adaptive thermal processing. Neural networks are used by Smith and Boning [81] and Card et al. [7] to develop empirical nonlinear control models. Card et al. uses a dynamic neural network to control a plasma etch process and Smith and Boning design an EWMA R2R controller that utilizes a neural network for self-tuning purposes. Most of the previously outlined methods do not explicitly model process disturbances or accommodate processes with constraints, and they do not address multiple input multiple output (MIMO) systems. To address these issues a model predictive control (MPC) strategy [60] from the traditional chemical industry is applied to R2R control [59]. Campbell [6] applies MPC to CMP processes and Edgar et al. [18] extends the MPC algorithm to lithography and rapid thermal processing reactors. The stability conditions and tuning guidelines for R2R controllers have also been examined. Ingolfsson and Sachs [39] examine the stability and the sensitivity of an EWMA controller and Good and Qin [30] analyze the stability of MIMO R2R controllers with metrology delay. An analysis of sample size effects on double EWMA controller stability is provided by Tseng and Hsu [91]. Methods for dealing with measurement and metrology delays and inconsistent data have also been developed. Wang et al. [93]

propose a method for handling metrology when time varying drifts are present and Smith and Boning [82] develop a method for implementing double EWMA with intermittent data sampling.

## 2.2 Module Level Control

R2R controllers have significantly improved manufacturing capabilities but opportunities for further improvement are missed because typical R2R controllers are set up to operate independently on individual processing steps without any knowledge of other manufacturing steps. Since manufacturing steps are often correlated, information from previous processing steps may be relevant to the current processing step and can be used to better control the current step. R2R controllers can be designed to allow feedforward and feedback information between processes to achieve better performance [17]. This type of control is also known as module level control in semiconductor manufacturing. These controllers take the CD measurement from a completed step and provide that data to the controller for the next step. At the next step the measurement is used to determine the recipe for the process that will keep the process as close to target as possible given the measurement information provided. Sachs et al. [77] first suggest using feedforward information in a R2R controller design. They propose adding the measurement from the previous step to the model used for R2R control and urge caution when using feedfoward control because of the possibility of increased instability. Stoddard [86], Ruegsegger et al. [74], and El Chemali et al. [20] also suggest applications of feedforward control for semiconductor manufacturing processes. Feedforward controller stability issue arises from the possibility of making adjustments based off of noisy measurements. If this occurs the output variance may end up greater than it was before feedforward control is applied and it can even make the process unstable. The work by Ruegsegger et al. addresses the potential stability problem that is present during feedforward control.

## 2.3 Multistep Control

R2R controllers with feedforward control improve the output of the processing step that receives the fedforward information but do not address coordinating multiple steps, via a multistep controller, with end-of-process measurements in mind. An example of this is the difference between feedforward control for a lithography and etch sequence and multistep control of the steps involved in a lithography process. In the lithography and etch sequence the lithography CD is fed forward to the etch process so that the etch process controller can compensate for variations in the lithography CD, but the lithography CD is not adjusted to better control the etch CD. In a three step lithography process a multistep controller can adjust the outputs of steps one and two to better control the CD measurement made after step three is complete. Only a few works have proposed multistage controllers that adjust the process inputs at multiple stages to control an end-of-process measurement. A general algorithm, that is not specific to semiconductor manufacturing, is proposed by Vaidyanathan [92]. It is a batch-wise myopic within batch optimal controller that optimizes the current batch across process steps and uses Bayesian techniques to account for uncertainty. A multistage controller for a three stage photolithography sequence is proposed by Leang et al. [47]. The controller determines input set points by working backwards from the target value for the last process step. As it works backwards inputs for each process are determined by centering each one into a predetermined acceptable region of operation for that particular process. The predetermined region of acceptable inputs is defined using a Monte Carlo simulation of the process that is performed using the empirical equations used for controlling the process. Kim and May [42] develop a neural network based multistage controller for a four step via formation process sequence. Neural network models for each of the individual processes and the overall process are developed. The inputs to the overall neural network model are the adjustable inputs to each of the individual processes and the outputs are the four outputs of the last individual processing step. A genetic algorithm is used to generate new recipes for the individual processes when an adjustment needs to be made. Another multistage algorithm for a photolithography process is presented by Fenner et al. [21]. The algorithm differs from the one developed by Leang et al. in that it allows for multiple goals between stages to be optimized. The algorithm also takes into account the synchronization of the process steps not considered in previous work, and uses data from past runs for handling uncertainty.

## 2.4 Fab-wide Control

The advantages of multistage control for sub-processes within a semiconductor manufacturing step can be extended to the whole manufacturing process. In this context the steps that are adjusted to control the end of process parameters are each of the manufacturing steps. The use of a fab-wide control strategy is first mentioned by Telfeyan et al. [90]. Telfeyan et al. acknowledge the need for a fab-wide controller and they design a CMP R2R controller for use in a fab-wide control strategy. A framework for fab-wide control, as well as supporting software and automation system requirements, is presented by Edgar et al. [17]. They recommend a computer integrated manufacturing system that would aid in data movement between equipment in the fab, as well as keep track of controller model and recipe data. Chaudhry et al. [8] propose the use of an active database to facilitate data availability and decision making for a multistep fab-wide controller. The works discussed so far do not specify the details of a fab-wide control scheme, such as deciding the important outputs to control or how to coordinate the processing steps. Qin and Sonderman [71] are the first to address important outputs to control by suggesting that the end-of-process measurement of interest in fab-wide control be the electrical parameters measured at the end of processing. Moyne [55] suggests that yield should be the focus of fab-wide control. Qin et al. [67] adds further detail to fab-wide controller design by proposing a three level hierarchical control framework that integrates the fab-wide controller with module level and R2R controllers to facilitate fab-wide control in terms of electrical parameters. Harrison et al. [34] are the first to present an algorithm for determining the control adjustments made between the processing steps in a fab-wide control scheme, as well as to address the modeling and process updating aspects of a fab-wide controller that are necessary for its functionality. They use a multi-step supervisory controller to control the erase time of a flash memory device and their work showes the potential variance reduction that can be achieved with a multi-step fab-wide control scheme. In work done by Qin et al. [68] the major components of a fab-wide algorithm are presented. Qin et al. also discuss the similarities and differences between fab-wide control and batch control in the chemical processing industry.

## 2.5 Complementary Literature

This section provides a review of other literature that is relevant to the work done in this dissertation.

### 2.5.1 Batch Control

Despite the differences outlined by Qin et al. [68], batch process control algorithms offer applicable methodologies for making adjustments at each step of a multi-step manufacturing process because most batch control algorithms manipulate the batch trajectory at specific points during the batch process. Each processing step in a multi-step process can be viewed as analogous to each segment of the batch trajectory. Data from completed processing steps in a multi-step manufacturing process is equivalent to batch trajectory data from the completed segments and the batch control algorithms can be modified to determine set points for future processing steps. Methods for batch trajectory manipulation using theoretical models are described by Crowley and Choi [11] and Ruppin et al. [75]. Both of these strategies use sequential quadratic programming techniques at specified time intervals but are difficult to implement because they require detailed theoretical model knowledge or are computationally intensive. In order to design batch controllers that can make use of historical data several controllers that make use of empirical models are proposed. Russell et al. [76] develop a controller for use with regression and neural network models and Flores-Cerrillo and MacGregor [22] present a controller for within batch control, as well as batch to batch control, that is based on a partial least squares model. The partial least squares model is used to represent the process with a reduced number of latent variables but the control adjustments are still made in the original variable space. In later works by Flores-Cerrillo and MacGregor [24, 25] control algorithms are proposed that perform the control moves in the latent variable space.

### 2.5.2 Real Time Optimization

An important feature of the proposed EPC scheme is the ability to provide real-time optimized targets to lower-level controllers. Real-time optimization is used widely in feedback and model predictive control for process optimization and optimal trajectory calculations. Mahadevan and Doyle [49] present efficient optimization approaches for nonlinear model predictive control and Pistikopoulos et al. [64] suggest the use of an off-line parametric tool for online optimization. Marlin and Hrymak [50] suggest real time operations optimization methods for continuous processes. Research has also focused on the model requirements and the accuracy necessary for real-time optimization. Yip and Marlin [102] analyze the effects of model fidelity on real-time optimization performance and Forbes et al. [26] discuss model requirements for good optimization performance. Real-time estimation has also been used for online parameter estimation and model updating. Yip and Marlin [101] propose a method for accommodating multiple data sets when performing model updates in real-time optimization and Krishnan et al. [43] discuss a robust parameter estimation method for online optimization.

### 2.5.3 Missing Data Algorithms

Missing data algorithms are methods for estimating missing process measurements by taking advantage of the information stored in the correlation structure of the historical data for the process and the currently available data. These algorithms are used when the process is modeled using latent variable methods, such as partial least squares or principle component analysis. The performance of the EPC algorithm presented in this dissertation is improved when combined with a missing data algorithm so a brief review of the literature is pertinent. There are two types of missing data - measurements that are missing in a historical data set that will be used for model building and measurements that are missing from a recently measured data vector when a model already exists. Algorithms for building latent variable models with missing data in the model building data set are proposed by Wold et al. (1983) and Marten and Naes [51, 97]. Nelson et al. [61] present three algorithms for predicting new missing data from an existing model and analyze the error associated with each of the methods.

## 2.6 Summary

EWMA based R2R controllers are the foundation of advanced process control in semiconductor manufacturing. Literature surrounding the development and advancement of R2R controllers, as well as the basic formulation for an EWMA based R2R controller, are reviewed in this chapter. Works describing the coordination of multiple manufacturing steps by using module level control and multistep control are also presented. The idea of coordinating multiple steps can be extended to fab-wide control and a literature review of the fab-wide controller literature is presented. Chapter 3 elaborates on the specific requirements for implementing a fab-wide control algorithm and later chapters will bring together fabwide control and the work done in batch control and missing data algorithms.

# Chapter 3

# Electrical Parameter Control Components and Requirements

The fab-wide controller developed in this dissertation focuses on electrical parameters as the end of processing parameters of interest. The proposed control scheme would provide a top-level controller that integrates independently operating R2R controllers and manufacturing steps without R2R control by providing them with optimized set points. The optimized set points are determined with the goal of reducing electrical parameter variation. This multi-step electrical parameter control (EPC) strategy is illustrated in Figure 3.1. The fab-wide controller uses metrology data from the completed processing steps, nominal or estimated values for the subsequent processing steps, and a device model to determine optimized targets for later processing steps. The electrical test measurements are used to update the model. This control scheme makes use of later processing steps to respond to errors introduced earlier in the manufacturing process. Process variations can be reduced by applying an EPC scheme that uses available data from completed processing steps to directly control electrical parameters. The direct control of electrical parameters via a fab-wide control framework introduces a new layer of control that can compensate for metrology drifts, coordinate R2R controllers by generating optimized targets for each processing step, and reduce variation in electrical parameters.

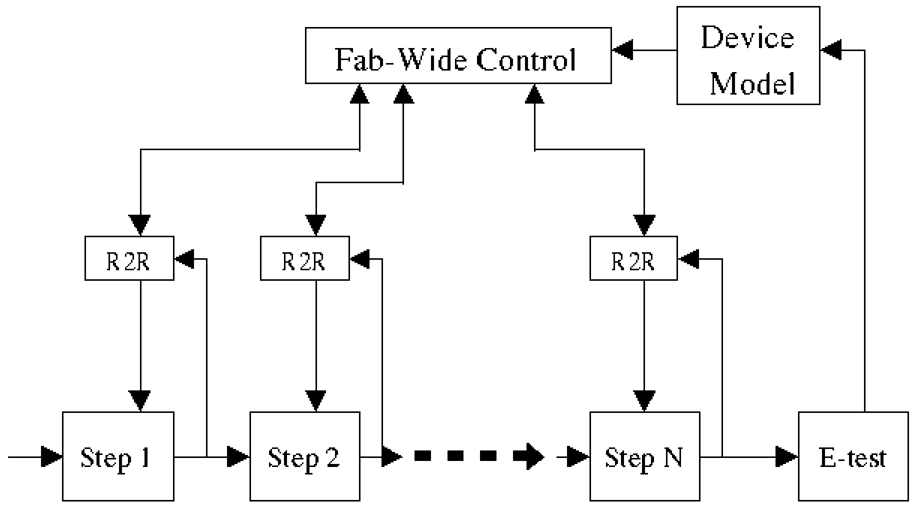


Figure 3.1: Fab-wide control block diagram

The EPC design involves three major components: a model used for prediction that relates processing parameters to electrical parameters, an algorithm to determine the optimal set points for subsequent processing steps using data from completed processing steps, and a method for updating the prediction model so that continues to represent the current process. The specific requirements for each of these components are expanded upon in the following sections.

## 3.1 Modeling

A prediction model relating geometric properties to electrical parameters for logic devices is necessary for electrical parameter control. A good prediction model is important to the success of the EPC controller because control moves are based off of the model. A poor model will lead to poor control and the possibility of increasing the output variation instead of reducing it. Most existing semiconductor process models relate operating conditions to geometric parameters, such as a model relating the etch CD to the etch time, which is a processing parameter that can be adjusting during etch operations. These types of models are primarily used for R2R control and each step in the manufacturing process has different models associated with that step. These models work well for R2R control but are not useful for predicting the electrical properties of completed wafers. The development of a model that can predict electrical parameters from metrology data is necessary for the EPC scheme. In addition to predictive capabilities, the model needs to generate predictions in real time because the model outputs need to be calculated before the next processing step can occur. One needs to be able to optimize the model in real time because it will be used in the determination of optimal set points for subsequent processing steps during manufacturing.

Both physics-based first principles models and data-driven models are able to meet the requirements needed for fab-wide control. There are many advantages and disadvantages for each type of model. Physics-based models are valid for large operating ranges and data-driven models are only useful within the range of data used to build the model. Very specific knowledge of the device being modeled, as well as enough data to sufficiently estimate all model parameters is required for physics-based models. Also, these models tend to be complex and are not good candidates for real time prediction and optimization. Also, most existing physics based models only predict one electrical parameter so multiple complex models would be needed to predict each of the electrical parameters of interest. For these reasons data driven models are used for this project. Data-driven models do not require as much device specific information, however knowledge of the process and device functionality is important for selecting the relevant inputs for the model. Of the approximately 300 processing steps involved in device fabrication only a subset of these processes are useful for controlling each electrical parameter. Also, data driven models make it easier to use different data sets for different types of devices, without having to find accurate physics-based models for each type of device.

Many types of data-driven models exist in the literature today. Some of the most widely used and simple to develop models are regression models [54, 95]. Regression techniques can be used to develop linear and nonlinear models and includes partial linear regression models which are also known as partial least squares (PLS) models. Other data driven modeling methods include neural networks [16, 73] and Bayesian models [10, 28]. Neural networks are an attempt to mathematically replicate the connectivity, operations, and pathway formation, via learning behavior, of neurons in the brain. Complex combinations of functions are laid out in a network structure and the network is trained to determine the weights associated with the various combinations of functions. Bayesian modeling is a method for using probability distributions of the data to determine distributions for model parameters, as opposed to point estimates. Bayesian model averaging is a technique for combining multiple models when it is unclear which model best describes the system [72]. Regression as well as Bayesian models are considered in EPC algorithm development.

## 3.2 EPC Algorithm

The EPC algorithm takes a device model of the device being manufactured, metrology data from completed processing steps, and expected values for the future processing steps and uses this information to predict the electrical parameter values for the wafers currently being processed. Based on how close the predicted electrical parameter values are to the target values a decision to make a control move is made. If a control move needs to be made the device model and available metrology data from completed steps are used to determine optimized targets for the subsequent processing steps. These targets are calculated to ensure that the wafers meet all of the electrical parameter requirements at the end of processing. The optimized targets are the set points for the lower-level R2R controllers and they are calculated by minimizing the difference between the predicted and targeted electrical parameter values. The objective function for the minimization is defined in Equation 3.1.

If no control move needs to be made the targets for the future processing steps remain at the nominal values. A flowchart of the EPC algorithm is shown in Figure 3.2.

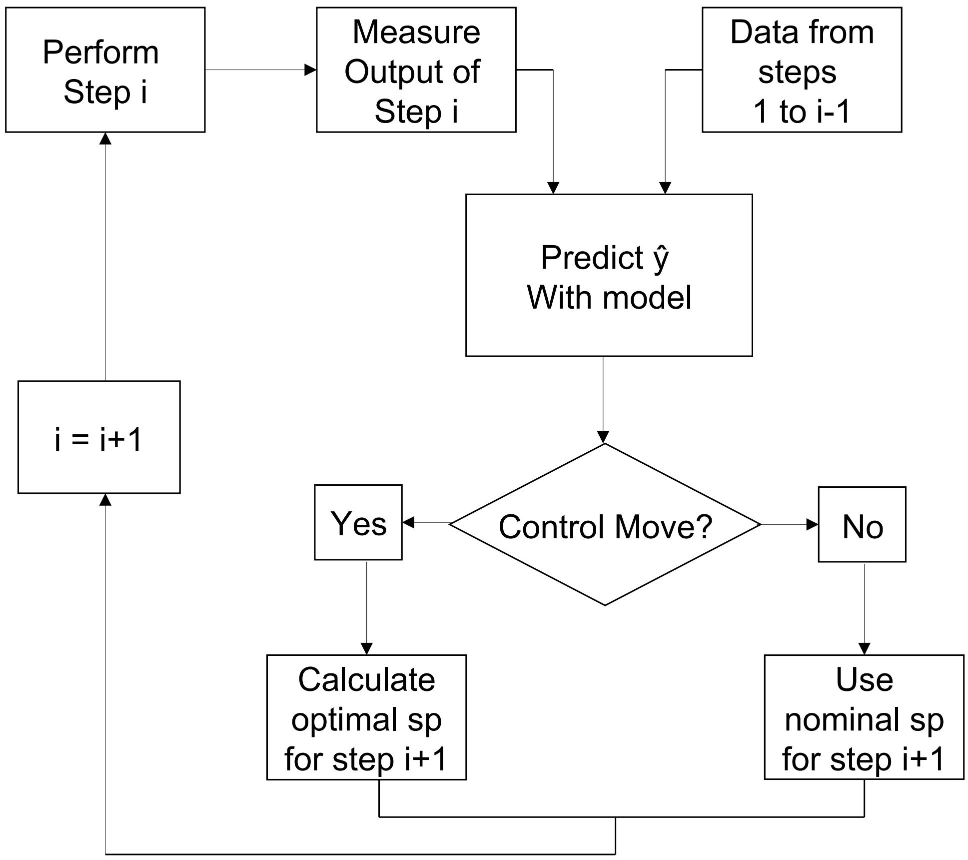


Figure 3.2: Flow chart of the EPC algorithm for step i

$$\min_{\hat{x}_2} \|y_{target} - f(x_1, \hat{x}_2)\|^2 + \lambda_{EPC}\|\hat{x}_2 - x_{2,nom}\|^2 \tag{3.1}$$

$$s.t. \quad x_{2,min} \leq \hat{x}_2 \leq x_{2,max}$$

$$y_{min} \leq f(x_1, \hat{x}_2) \leq y_{max}$$

Here vector $y_{target}$ contains the desired electrical parameter values, $x_1$ represents the input values for completed processing steps, $\hat{x}_2$ represents the input values for subsequent processing steps, and $\lambda_{EPC}$ is a weighting factor. The constraints set the appropriate lower and upper limits for $\hat{x}_2$. The variable being optimized is $\hat{x}_2$. The predicted electrical parameter values are obtained from the function, $f$, which is a model relating electrical parameters to input parameters, $x_1$ and $\hat{x}_2$. The second term in the objective function penalizes large changes in the input values, so that the input targets will not change drastically between groups of wafers that are processed. This helps produce stable targets for the lower-level R2R controllers. The constraints and target values can be specified by product specifications and requirements or they can be user determined values that are reasonable for the process. If $f$ is linear then the objective function is quadratic so a minimum is guaranteed. If $f$ is not linear then more analysis is necessary to determine if a minimum exists and if it is a global or a local minimum.

The formulation of the objective function is flexible and weights can be added in the event that some of the electrical parameters are more important than others. Also, other functions of $\hat{x}_2$ can be added to the objective function if necessary. This is useful if there are other processing parameters that are functions of $\hat{x}_2$ that need to be kept within specifications but they do not affect the electrical parameter values.

The target re-optimization occurs after new metrology data is measured for each processing step so only the most recent optimized target gets used.

It is not necessary to calculate optimized targets for all of the subsequent processing steps but it allows the adjustments that are made at each step to be smaller because no one step is attempting to bring the outputs back on target. The optimization needs to be performed in a timely manner since the wafers can not start the next processing step until the optimal set point for that step has been determined.

## 3.3 Model Updating

As metrology data, processes, and materials change over time the model needs to be updated to reflect these changes. The model updating algorithm will depend on the type of model used for the EPC scheme. For a simple linear model updating will be accomplished with model parameter estimation from recent process data. The model parameters estimates are obtained by minimizing the difference between electrical test data and predicted values for the electrical tests. The least-squares objective is the dual of the EPC objective

$$\min_{\theta} \sum_{k=1}^{N} \|(y(k) - f(x_1(k), \theta)\|^2 + \lambda_{MU} \|\theta - \theta_{nom}\|^2. \tag{3.2}$$

Where $y(k)$ is a vector of electrical parameters, $f$ is the model that predicts electrical parameters from the inputs, $x_1$ is the input vector, $\theta$ is a vector of the model parameters to be estimated, $\theta_{nom}$ is the current vector of model parameters, $\lambda_{MU}$ is a weighting factor, and $N$ is the number of lots or wafers that have completed processing since the model was last updated. The second term in the minimization penalizes large changes in model parameters between updates. This ensures that the model is updated in response to long term changes in the operating conditions but not to transient errors.

For a PLS model, updating can be accomplished by using a recursive partial least squares (RPLS) algorithm with a forgetting factor [66] or adding a model bias term and updating the model bias using an exponentially weighted average, similar to that used in R2R control. Both methods allow the PLS model to be updated without the need to invert the data matrix, which can be time consuming if the data set is large. The forgetting factor enables the RPLS algorithm to adjust to slowly time varying processes by lowering the weighting factor for older data. This is accomplished through the weighted average in the model bias update.

## 3.4 Electrical Test Measurement Delay

Due to the nature of semiconductor manufacturing there is significant delay time between wafer processing and wafer electrical test measurements. The delay time can be anywhere from a few weeks to three months, depending on the specifics of the manufacturing process. Electrical measurements are often made after each metallization process before another layer of interconnects is patterned onto the wafer. These measurements can be used in lieu of the end of process electrical parameter measurements to reduce the delay time between processing and measurement. Delays this large can make modeling and controlling the process difficult. It may seem unreasonable to make process adjustments when the impact of the adjustments will not be known for weeks at a time, but this amount of process lag time is present when manual adjustments are made by process engineers in the fab, even when no EPC controller is present. Instead of having process engineers review data and make adjustments when the electrical parameters drift too far away from target the EPC controller can make automatic adjustments before the process drifts significantly. In order to have confidence in the adjustments that are being made it is necessary to have a model that accurately reflects the process under control. That is why developing and updating the EPC model is instrumental in applying EPC effectively. The ability to constrain the process adjustments also allows control moves to be made with more confidence.

## 3.5 Summary

The requirements for the device model and a basic mathematical formulation of the EPC algorithm and the model updating algorithm are described in detail. There are many different ways to implement an electrical parameter controller and Chapter 4, Chapter 5, and Chapter 6 describe three implementation options in detail. Simulations are used to determine the effectiveness of each implementation and the limitations of each are discussed.

# Chapter 4

# Least Squares EPC

The first step in evaluating the effectiveness and potential of the EPC algorithm is to test it on a simple system using a basic implementation. The more complex aspects of the algorithm can be developed using the simple implementation as a starting point. The results from this system can be used as a baseline for evaluating the algorithm when it is applied to more complex manufacturing scenarios.

## 4.1 Modeling

In this implementation the predictive model used to relate electrical parameters to processing parameters is a simple linear model. It can accommodate multiple inputs as well as multiple outputs and the parameters can be determined from a least squares solution using a reasonable sized data set and assuming all of the inputs are independent. The model equation and least squares parameter regression solution are given in Equation 4.1 and Equation 4.2,

$$Y = \theta X + \varepsilon \tag{4.1}$$

$$\theta = (X^T X)^{-1} X^T Y. \tag{4.2}$$

Where $X$ is the $n \times 1$ input matrix, $Y$ is the $m \times 1$ output vector, $\varepsilon$ is the model residual vector, and $\theta$ is the $m \times n$ regressed parameter vector. To estimate the model parameters $\theta$, $X$ and $Y$ data from experimental design and historical operations can be used to obtain the least squares estimates.

## 4.2 EPC algorithm

The EPC algorithm uses data from completed processing steps, the linear model described in Equation 4.1, and expected values for future processing steps to predict electrical parameter values. The predictions are used to determine if a control move needs to be made. If so, the model and the available data for completed steps are used to determine the optimal set points for the subsequent processing steps. The optimal set points, $\hat{x_2}$, are calculated by minimizing the difference between predicted and targeted electrical parameters. The objective function described in Section 3.2 is modified using the linear model as the predictive function. The objective function specific to this implementation takes the form:

$$\min_{\hat{x_2}} \|y_{target} - \theta[x_1 \quad \hat{x_2}]\|^2 + \lambda_{EPC}\|\hat{x_2} - x_{2,nom}\|^2 \tag{4.3}$$

$$s.t. \quad x_{2,min} \leq \hat{x_2} \leq x_{2,max}$$

$$y_{min} \leq \theta[x_1 \quad \hat{x_2}] \leq y_{max}$$

The $x$ matrix gets partitioned into sub-matrices $x_1$, that contains data from the completed processing steps, and $\hat{x_2}$ that represents the optimal set points for the subsequent processing steps. In order to rearrange the objective function into the standard quadratic form the parameter matrix, $\theta$, needs to be partitioned in a similar fashion as the $x$ matrix. The parameters that multiply $x_1$ are partitioned into $\theta_1$ and the parameters that multiply $\hat{x_2}$ are partitioned into $\theta_2$. Once this is completed the objective function in Equation 4.3 can be rearranged into the standard quadratic programming form given in formulation 4.4

$$\begin{aligned} \min_{\hat{x_2}} \frac{1}{2}\hat{x_2^T} H \hat{x_2} + G^T \hat{x_2} \\ H = \frac{1}{2}(\theta_2^T \theta_2 + \lambda_{EPC} I) \\ G = -((y_{target} - \theta_1 x_1)^T \theta_2 + \lambda_{EPC} x_{2,nom}^T). \end{aligned} \tag{4.4}$$

The unconstrained case has a closed form solution for the optimal set points given in Equation 4.5.

$$\hat{x_2} = -(2H)^{-1} G^T \tag{4.5}$$

The constrained case requires using a quadratic solver such as the quadprog function found in the Matlab Optimization Toolbox [37].

## 4.3 Model Updating

### 4.3.1 Model Bias Updating

The predictive model needs to be updated occasionally so that it accurately reflects the current process and can reject disturbances, such as drift. For this implementation the predictive model will be updated with a model bias term, $b_k$. The model bias term is a weighted sum of the difference between the predicted and actual electrical parameter values and the previous value of the bias term.

$$\hat{y}_k = \theta x_k + b_k \tag{4.6}$$

where

$$b_k = \lambda_{MU} b_{k-1} + (1 - \lambda_{MU}) \frac{1}{w} \sum_{i=1}^{w} (y_i - \hat{y}_i) \tag{4.7}$$

Here $w$ is the number of lots that have completed processing since the last model bias coefficient was calculated and $k$ is the number of times the model has been updated. Equation 4.7 is essentially an exponentially weighted moving average filter.

### 4.3.2 Whole Model Updating

In some situations it may be desirable to update the model gain instead of just the model bias term. This is easily accomplished by using the new process data to recalculate the model parameters using Equation 4.2. Though the calculation is simple it is necessary to examine the validity of the parameter estimates when using data obtained with the controller turned on.

First a simple two input one output process is examined. In this simplified scenario the first process step is not adjusted and the target for the second processing step is determined using the measurement from the first processing step. The set point value for the second step can be calculated explicitly and is shown in Equation 4.8.

$$x_{2,sp} = \frac{y_{target} - \theta_1 x_1}{\theta_2} \tag{4.8}$$

Assuming the actual measurement for the second processing step is modeled as the set point value plus a measurement error, $\varepsilon$, the closed loop input data matrix, $X$, is populated as follows.

$$X = \begin{bmatrix} x_{11} & \frac{y_{target}-\theta_1 x_{11}+\varepsilon_1}{\theta_2} \\ x_{21} & \frac{y_{target}-\theta_1 x_{21}+\varepsilon_2}{\theta_2} \\ . & . \\ . & . \\ . & . \\ x_{n1} & \frac{y_{target}-\theta_1 x_{n1}+\varepsilon_n}{\theta_2} \end{bmatrix} \tag{4.9}$$

When the least squares model is recalculated this data matrix is used for the model parameter calculation in Equation 4.2. Since the second column is not a linear combination of the first column, the matrix, $(X^T X)^{-1}$, will exist and the estimates obtained with controlled data are valid. A similar analysis was done with a three input one output system where the set points for the future steps are calculated using the objective function in Equation 4.4. The second and third column of the resulting data matrix are shown in Equation 4.10 and Equation 4.11 respectively.

$$X_{col2} = \begin{bmatrix} (2H(\theta_2,\theta_3))^{-1}[(Y_T - \theta_1 x_{11})^T[\theta_2 \quad \theta_3] + \lambda[x_{2,nom} x_{3,nom}]]^T \\ . \\ . \\ . \\ (2H(\theta_2,\theta_3))^{-1}[(Y_T - \theta_1 x_{n1})^T[\theta_2 \quad \theta_3] + \lambda[x_{2,nom} x_{3,nom}]]^T \end{bmatrix} \tag{4.10}$$

$$X_{col3} = \begin{bmatrix} (2H(\theta_3))^{-1}[(Y_T - \theta_1 x_{11} - \theta_2 x_{12})^T \theta_3 + \lambda x_{3,nom}]^T \\ . \\ . \\ . \\ (2H(\theta_3))^{-1}[(Y_T - \theta_1 x_{n1} - \theta_2 x_{n2})^T \theta_3 + \lambda x_{3,nom}]^T \end{bmatrix} \tag{4.11}$$

As in the two input case, the second column is not a linear combination of the first column and third column is not a linear combination of the first and second columns so the data matrix, $(X^T X)$, will be invertible. This analysis shows that parameter updates can be performed with data taken when the controller is turned on.

The ability to recalculate the least squares model with closed loop data does not necessarily guarantee that the estimates will reflect the actual process gain. In order to determine whether using closed loop data affects the quality of the gain estimates a simulation is performed. The simulation assumes there are ten inputs and three outputs and control moves are made on steps three through ten. Three of the inputs are independent and the other seven are correlated with the three independent inputs. The first 500 data points

Table 4.1: Output y1% difference between actual model gains and the model gain used for EPC control, the model gain calculated from data with no control, and the model gain calculated from closed loop data with EPC control

| | | y1 | |
|---|---|---|---|
| Input | EPC Model | Open Loop | Closed Loop |
| x1 | 3.51% | 3.16% | 0.68% |
| x2 | 11.80% | 27.52% | 17.96% |
| x3 | 2.04% | 6.50% | 5.98% |
| x4 | 26.95% | 17.70% | 18.40% |
| x5 | 1.36% | 0.55% | 1.69% |
| x6 | 52.82% | 53.46% | 58.44% |
| x7 | 0.80% | 8.11% | 5.37% |
| x8 | 4.59% | 1.53% | 0.47% |
| x9 | 7.01% | 3.24% | 3.27% |
| x10 | 4.45% | 0.45% | 5.87% |

are used to build the least squares model used in the EPC algorithm. The simulation is run with no control and with the EPC controller turned on. The second 500 data points are measured for each case. Table 4.1, Table 4.2, and Table 4.3 show the percentage difference between the actual model gains and the least squares estimates of the model gains used for EPC control, the least squares estimates of the model gains with no control, and the least squares estimates of the model gains with EPC control for each of the three outputs. The data indicates that using the closed loop data obtains estimates that are similar in quality to open loop estimates. Some of the variables, such as $x6$, show large errors in all estimates and this is because not all of the variables are independent.

Table 4.2: Output y2% difference between actual model gains and the model gain used for EPC control, the model gain calculated from data with no control, and the model gain calculated from closed loop data with EPC control

| | | y2 | |
|---|---|---|---|
| Input | EPC Model | Open Loop | Closed Loop |
| x1 | 15.05% | 10.41% | 14.45% |
| x2 | 11.26% | 29.56% | 30.20% |
| x3 | 5.88% | 0.88% | 1.23% |
| x4 | 2.44% | 6.58% | 7.52% |
| x5 | 4.96% | 0.76% | 2.69% |
| x6 | 1.96% | 6.17% | 5.50% |
| x7 | 9.58% | 10.01% | 1.44% |
| x8 | 8.93% | 2.89% | 2.77% |
| x9 | 20.26% | 22.25% | 10.93% |
| x10 | 5.44% | 20.34% | 1.28% |

## 4.4 Simulation

To evaluate the effectiveness of the proposed EPC algorithm a semiconductor fabrication process was simulated. Due to the importance of having an adequate model for EPC, a simulation to assess the effect of model quality on control performance was investigated, as well as the controller performance with various noise and constraint combinations. The simulated fab process has 13 processing steps and 4 outputs. The simulation is a two layer process where data for the 13 processing steps are generated using a R2R control simulation and data for the four electrical parameters is generated in a fab-wide simula-

Table 4.3: Output y3% difference between actual model gains and the model gain used for EPC control, the model gain calculated from data with no control, and the model gain calculated from closed loop data with EPC control

| | | y3 | |
|---|---|---|---|
| Input | EPC Model | Open Loop | Closed Loop |
| x1 | 1.91% | 47.81% | 30.95% |
| x2 | 18.03% | 5.11% | 14.95% |
| x3 | 16.00% | 11.80% | 10.56% |
| x4 | 4.50% | 2.28% | 5.46% |
| x5 | 5.72% | 11.25% | 5.10% |
| x6 | 9.81% | 5.48% | 2.31% |
| x7 | 6.12% | 4.91% | 1.65% |
| x8 | 94.51% | 45.31% | 17.38% |
| x9 | 1.85% | 19.00% | 14.06% |
| x10 | 11.43% | 0.17% | 2.72% |

tion using the R2R simulation data. A block diagram showing the simulation set up is shown in Figure 4.1.

### 4.4.1 R2R Controller Simulation

The R2R controller simulation inputs a set point for the output associated with the process the R2R controller is acting on and returns an output value for that process. The simulation includes an integrated moving average (IMA(1,1)) disturbance, metrology error, and R2R controller model error. IMA(1,1) noise is defined in Equation 4.12. IMA(1,1) is a nonstationary drift model that is applicable to manufacturing processes where machine parts degrade over time [3, 13]. The IMA disturbance is $d$, $\varepsilon$ is normally distributed noise with standard deviation $\sigma$ that is independent between wafers processed, and $\theta_{ima}$ is a parameter that represents the degree of nonstationarity.

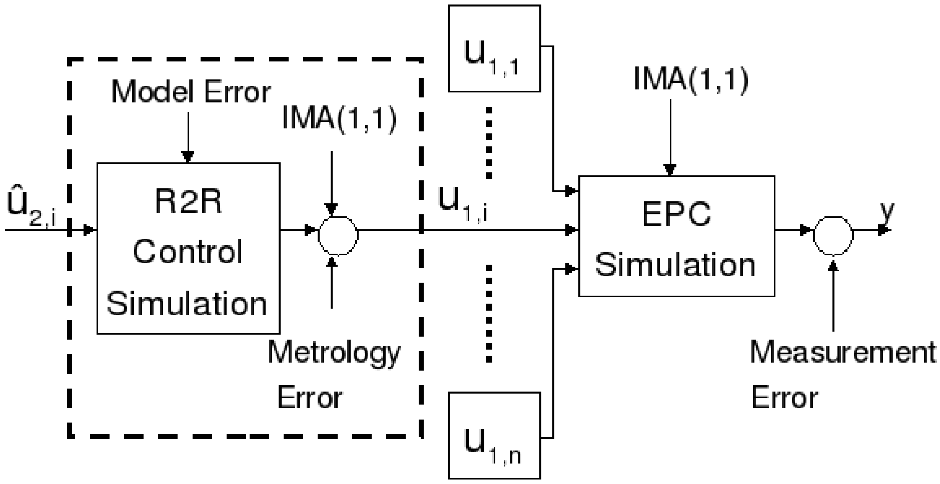


Figure 4.1: Block diagram of simulation data and noise

$$\Delta d_k = d_k - d_{k-1} = \varepsilon_k - \theta_{ima}\varepsilon_{k-1} \tag{4.12}$$

The simulation generates the output of the process according to Equation 4.13.

$$x_{cd,k} = b(1 + b_{error}\varepsilon_{b,k})u_k + a_k + d_k + m_{error}\varepsilon_{m,k} \tag{4.13}$$

The output generated by the simulation is $x$, $b$ is the model gain, $a$ is the model bias, $u$ is the process input that is defined by the R2R control law in Equation 2.2 and Equation 2.3, $d$ is the IMA disturbance, the model error is a normally distributed random error given by $\varepsilon_{b,k}$ with standard deviation $b_{error}$, and the metrology error is also normally distributed with standard deviation $m_{error}$. The values of the error parameters are listed in Table 4.4

Table 4.4: R2R Controller Simulation Errors

| Type | Error |
|---|---|
| Metrology Error, $m_{error}$ | 0.015 |
| R2R Controller Parameter Error, $b_{error}$ | 0.02 |
| IMA Standard Deviation, $\sigma$ | 0.005 |
| IMA Nonstationarity Parameter, $\theta_{ima}$ | 0.3 |

### 4.4.2 Fab-wide Simulation

The fabrication simulation generates electrical parameter data using a linear model relating the 13 inputs, which are the outputs of the R2R controller simulation, to the four electrical parameter outputs. The 13 inputs are independent of each other. Normally distributed measurement error and an IMA(1,1) disturbance are added to the electrical parameters. The electrical parameter values are generated from Equation 4.14.

$$y = \theta x + d + m_{error}\varepsilon_{ep} \tag{4.14}$$

Matrix $y$ contains the electrical parameters, $\theta$ is the gain matrix, $x$ are the outputs of the R2R controller simulations, $d$ is an IMA disturbance, and $m_{error}$ is the measurement error. The values for the error parameters are listed in Table 4.5.

The EPC algorithm requires a process model, so the simulation is run

Table 4.5: Fab-wide Simulation Errors

| Type | Error |
|---|---|
| Measurement Error, $m_{error}$ | 0.1 |
| IMA Standard Deviation, $\sigma$ | 0.1 |
| IMA Nonstationarity Parameter, $\theta_{ima}$ | 0.3 |

without the controller turned on to generate a training data set. A process model is derived from the training data set as described in Section 4.1. This ensures that the model used for simulating the process and the model used for control are not the same.

Once the predictive model is calculated the simulation is run with the EPC controller turned on. The EPC controller acts on one wafer at a time and uses all of the available CD measurements from the completed steps and the expected values of the CD measurements from the subsequent processing steps to predict the electrical parameters. The decision to make a control move is made if all of the predicted electrical parameter values are greater than one standard deviation away from their targets, based on the historical data used to build the model. If the decision to make a control move is made the optimal set points for the subsequent processing steps are calculated from Equation 4.4. The wafer then moves onto the next processing step and the optimal set point is used at the next step. In a real fabrication process adjustments can not be made at every step so in this simulation, adjustments are made at all processing steps except the first and the fourth steps.

## 4.5 Results

### 4.5.1 Model Quality Study

The predictive model is an important piece of the EPC algorithm because the control move decisions and the control moves are only useful if the model adequately represents the process that is being controlled. In order to get a better understanding of the model fidelity requirements to achieve adequate control a model quality study is performed. To demonstrate the effects of model quality on controller performance three test cases with models of varying quality were simulated. For these test cases there are no constraints on the controller adjustments and the only process noises are metrology error and an IMA disturbance. The model quality is evaluated using the coefficient of determination, $R^2$, which is well suited for evaluating regression models. The model's $R^2$ statistics were adjusted by changing the standard deviations of the normally distributed noise used in the electrical parameter IMA(1,1) disturbance and the electrical parameter measurement error. The $R^2$ statistic for each model is summarized in Table 4.6.

The simulation was run 100 times and the mean and standard deviation values for the four electrical parameter outputs, Y1, Y2, Y3, and Y4 were averaged. This study indicates that the model quality significantly impacts the algorithms effectiveness but even when using models with $R^2$ values on the order of 0.6 - 0.7 output variations can be reduced.

Table 4.6: Model quality effects on controller performance

| | Output | $R^2$ Value | STD Improvement |
|---|---|---|---|
| Case 1 | Y1 | 0.63 | 21.9 % |
| | Y2 | 0.64 | 21.2 % |
| | Y3 | 0.70 | 26.5 % |
| | Y4 | 0.73 | 28.5 % |
| Case 2 | Y1 | 0.81 | 43.0 % |
| | Y2 | 0.82 | 43.6 % |
| | Y3 | 0.85 | 49.6 % |
| | Y4 | 0.87 | 49.0 % |
| Case 3 | Y1 | 0.99 | 74.5 % |
| | Y2 | 0.99 | 75.4 % |
| | Y3 | 0.99 | 78.7 % |
| | Y4 | 0.99 | 79.3 % |

### 4.5.2 EPC Performance Study

Simulations are used to demonstrate the effectiveness of the EPC algorithm in varying noise and disturbance environments with and without constraints. Models with low $R^2$ values are used since it is difficult to obtain industrial data with better correlations. The model correlations are shown in Table 4.7.

The simulation used to evaluate the EPC algorithm performance with offset disturbances and constraints examines four test cases. Case 1 shows performance in the presence of the IMA(1,1) disturbance and process noise, Case 2 shows performance in the presence of the IMA(1,1) disturbance, process noise, and a metrology offset of 0.008 in the second step and -0.008 in the fourth processing step, Case 3 is the same as Case 1 except that the controller constrains the calculated set point value and the difference between the last set point and the calculated set point value, $\Delta\hat{u}_k = \hat{u}_k - \hat{u}_{k-1}$, and Case 4 applies the same constraints to Case 2. The constraints for Cases 3 and 4 are

$$-0.035 \leq \hat{u}_k \leq 0.035$$

$$-0.02 \leq \Delta\hat{u}_k \leq 0.02$$

The simulation was run 100 times and the mean and standard deviation values for the electrical parameters, Y1, Y2, Y3, and Y4 were averaged. A summary of this data is shown in Table 4.7. Figures 3-6 show a sample output for each case. The first 100 data points are the training data set and the second 100 data points show the improvement after the controller has been turned on. For all four cases the EPC algorithm keeps the outputs very close to their set point values of zero and significantly reduced variations in the outputs as shown by the reduction in standard deviation values. As expected the controller is less effective when constraints are in place as indicated by less improvement in the standard deviation for Cases 3 and 4 as compared to Cases 1 and 2 respectively.

## 4.6 Industrial Case Study

Industrial data obtained from Texas Instruments is used to further demonstrate the effectiveness of the EPC algorithm. The data set used consists of 13 input variables, such as gate lengths and doping concentrations, and four outputs variables, such as parametric transistor drive current and leakage current. The data is taken when no inline set points are manipulated with a supervisory controller and the data is modeled using ordinary least squares.

Table 4.7: Output mean and standard deviation improvements with control

| | | $R^2$ | Mean | STD Improvement |
|---|---|---|---|---|
| Case 1 | Y1 | 0.63 | -0.039 | 21.9 % |
| | Y2 | 0.64 | -0.034 | 21.2 % |
| | Y3 | 0.70 | -0.043 | 26.5 % |
| | Y4 | 0.73 | -0.031 | 28.5 % |
| Case 2 | Y1 | 0.63 | 0.032 | 17.6 % |
| | Y2 | 0.64 | 0.029 | 21.0 % |
| | Y3 | 0.70 | 0.0086 | 24.1 % |
| | Y4 | 0.73 | 0.026 | 27.1 % |
| Case 3 | Y1 | 0.63 | -0.042 | 9.0 % |
| | Y2 | 0.64 | -.04 | 12.0 % |
| | Y3 | 0.70 | -0.053 | 14.1 % |
| | Y4 | 0.73 | -0.023 | 16.7 % |
| Case 4 | Y1 | 0.63 | -0.047 | 10.8 % |
| | Y2 | 0.64 | -0.052 | 13.8 % |
| | Y3 | 0.70 | -0.052 | 15.8 % |
| | Y4 | 0.73 | -0.028 | 19.8 % |

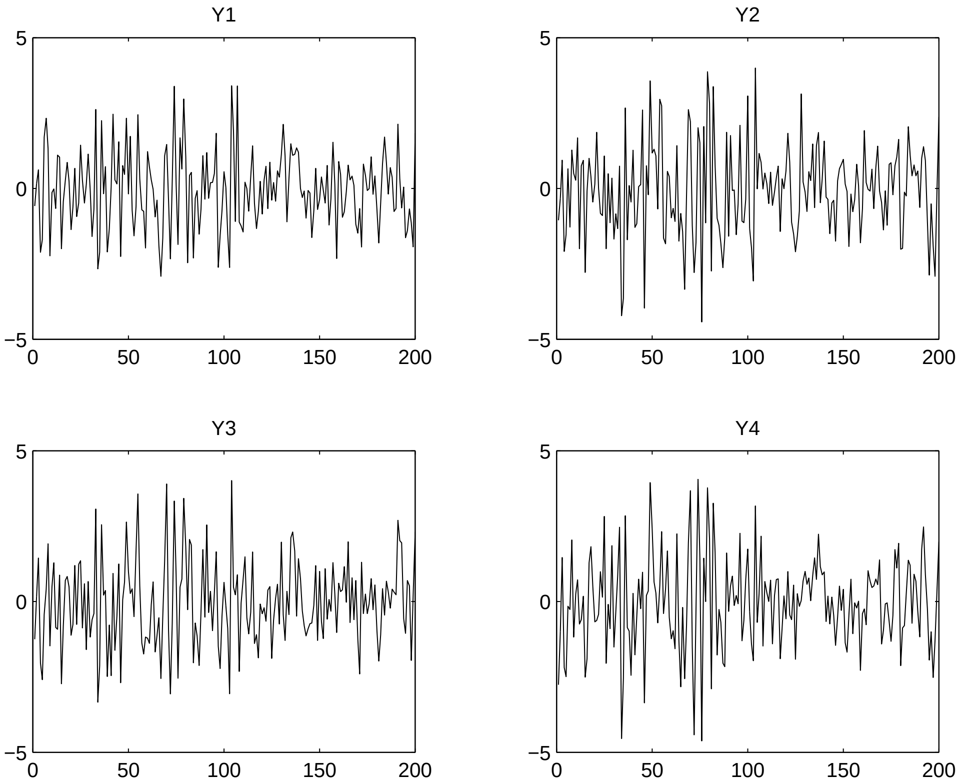


Figure 4.2: Case 1 - IMA(1,1) disturbance and process noise

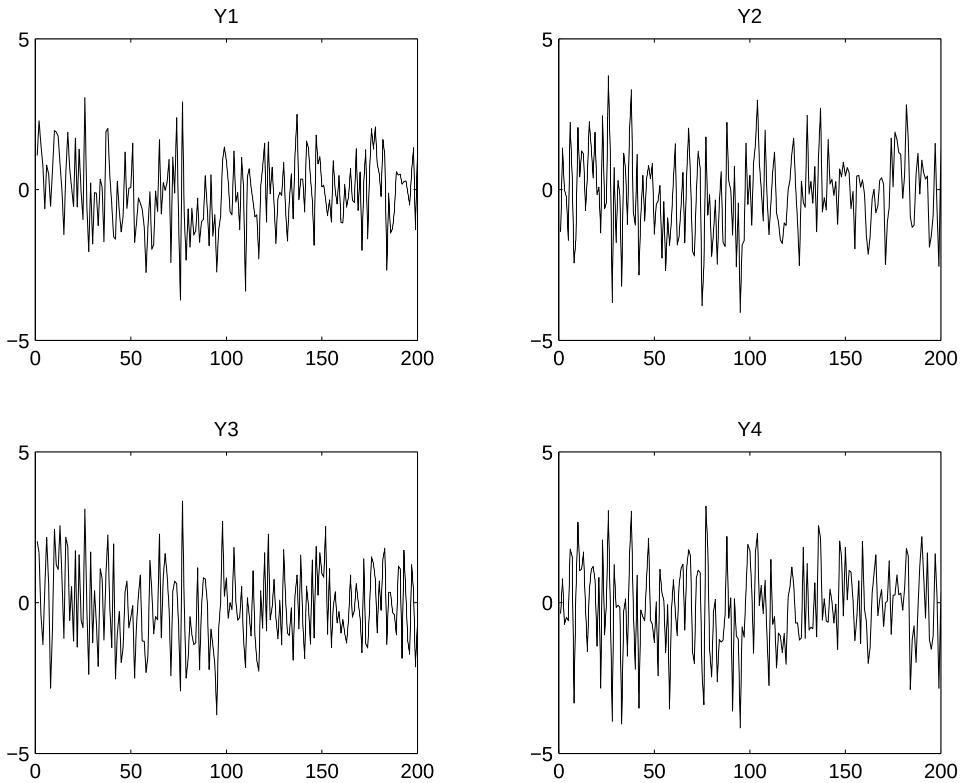


Figure 4.3: Case 2 - IMA(1,1) disturbance, process noise, and metrology offset at Steps 2 and 4

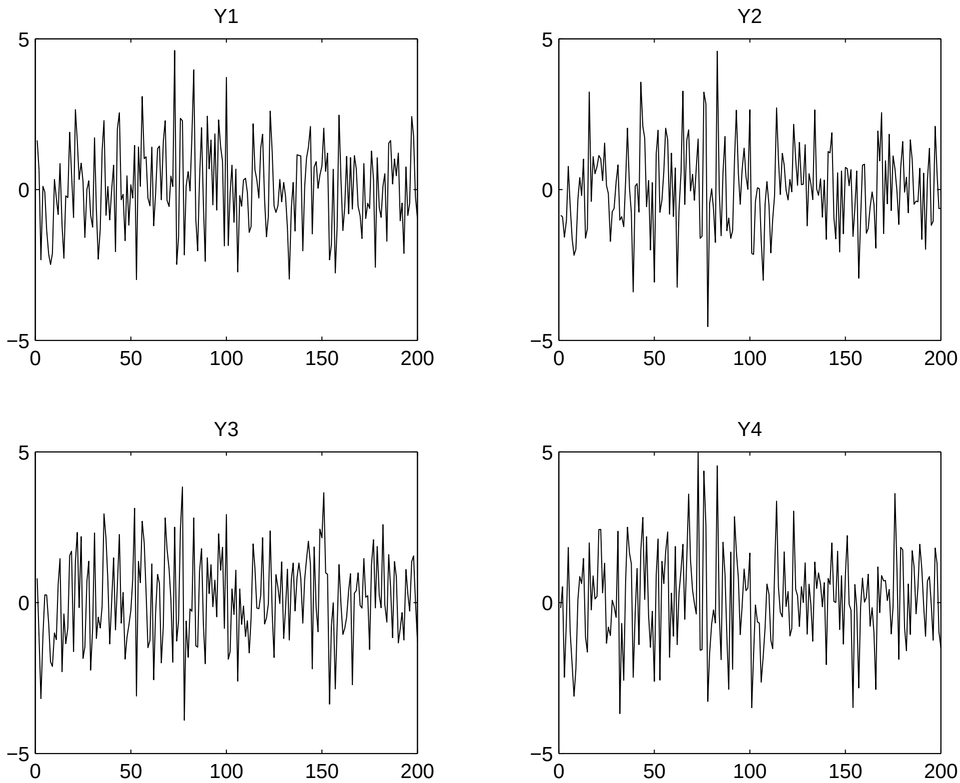


Figure 4.4: Case 3 - IMA(1,1) disturbance and process noise with constraints on the manipulated variable

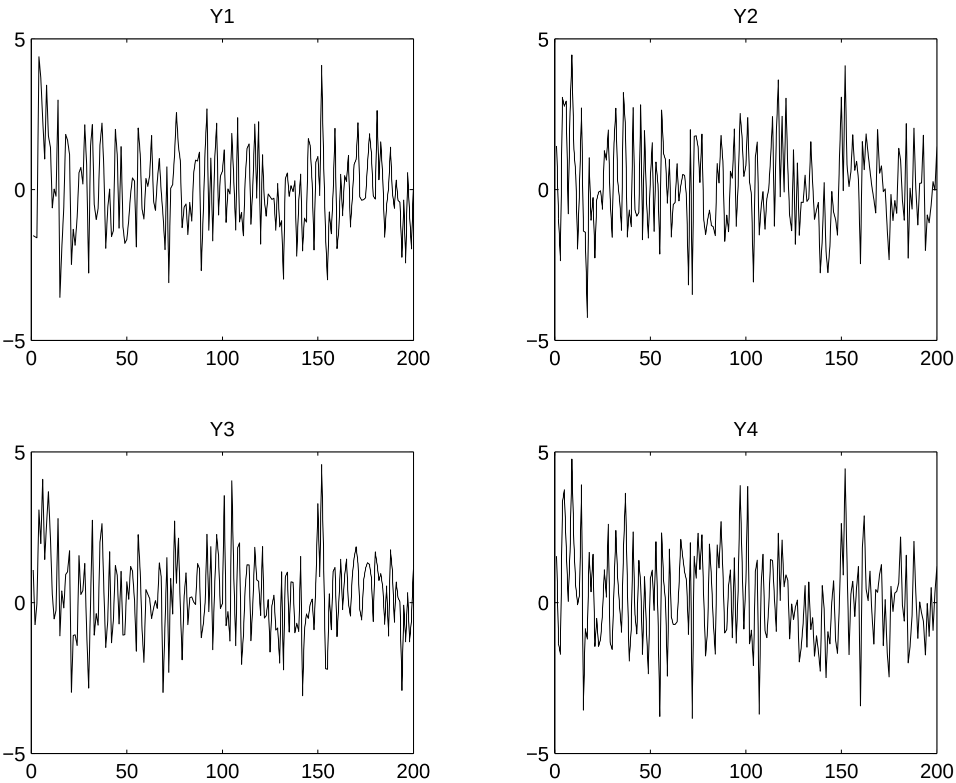


Figure 4.5: Case 4 - IMA(1,1) disturbance, process noise, and metrology offset at Steps 2 and 4 with constraints on the manipulated variable

The change in the process output values with EPC applied are calculated by taking the difference between the actual set points and the optimized set points calculated by the controller and multiplying them by the corresponding coefficient in the least squares model as shown in Equation 4.15.

$$y_{cntrl} = y_{actual} + \theta_x (x_{cntrl} - x_{actual}) \tag{4.15}$$

For this particular process the controller was only allowed to determine the set point for the 13th step of the process. Results are summarized in Table 4.8. Figure 4.6 shows the N channel leakage current values before and after control is applied. The horizontal lines indicate the standard deviation before and after control is applied. The data shows that the EPC algorithm effectively reduces the output variation in all four process outputs.

This industrial example is still an idealized case since real production fabs have multiple products running on multiple tools and metrology delay impedes the access to real time information. In a real manufacturing environment it may be necessary to use different models for different tools because the offsets and biases included in the model might be different. If a processing step has metrology delay the process target or average value can used for prediction purposed until the data becomes available.

Table 4.8: Industrial Data Improvements

| Output | STD no control | STD with EPC |
|---|---|---|
| Y1 | 1.0 | 0.70 |
| Y2 | 1.0 | 0.63 |
| Y3 | 1.0 | 0.77 |
| Y4 | 1.0 | 0.74 |

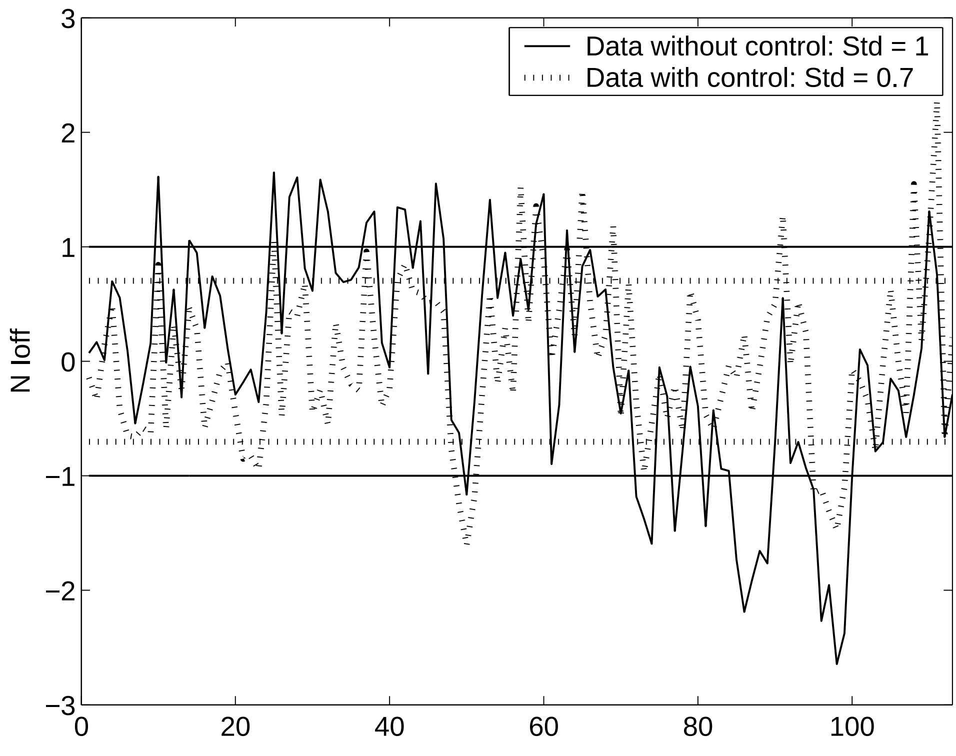


Figure 4.6: Industrial Data with and without control

## 4.7 Limitations

The EPC algorithm described in this chapter has many limitations based on the assumptions made about the manufacturing process that is being controlled. The assumptions made in this formulation are that all of the CD variables are independent, CD metrology data is available for each wafer before the next processing step is performed, and adjustments can be made on a wafer to wafer basis. In most manufacturing environments these assumptions will not hold. Process outputs from step to step are usually correlated, which means that an adjustment made in step $n$ can also affect the outcome of step $n+1$. Those effects were not accounted for when determining set points for subsequent processing steps. Data availability is another important problem that was not addressed in this formulation. In most semiconductor fabrication processes CD metrology data is not measured for every wafer in every lot. Instead, a subset of wafers in each lot is measured and the same wafer is not necessarily measured after each processing step. This provides challenges for model building as well as process adjustments. Data availability is also affected by metrology delay. In these situations metrology from the most recent step is not available before the next lot begins processing on that step. The next chapter presents enhancements to the EPC algorithm in order to address these limitations.

# Chapter 5

# Latent Variable EPC

The usefulness of the EPC algorithm depends on its ability to provide variance reduction in real manufacturing environments. As mentioned in Section 4.7, there are many complex issues that need to be addressed for this to happen. The major problems are correlations between the CD data that need to be taken into account when adjustments are made, the unavailability of data for each wafer after each processing step, which affects model building as well as control, and the inability to make wafer level adjustments at each processing step.

In order to take the correlation structure of the data into account a PLS model can be used instead of a least squares model. The PLS model can be combined with a missing data algorithm so that models can be built from data sets where measurements are not available for each wafer from each processing step. This greatly increases the amount of historical data that can be used for model building. The PLS model and a missing data algorithm can also be used to get better electrical parameter predictions. The electrical parameter values are predicted using the measured data from completed processing steps and historical or nominal values for the subsequent processing steps. Then a control decision is made based on the prediction. Instead of using the nominal or historical average values for the subsequent processing steps the missing data algorithm can be used to predict more accurate values. This way better electrical parameter predictions are made that lead to better control decisions. Also, the PLS model and the missing data algorithm can be used to compensate for metrology delay. The metrology delayed measurements can be replaced with values calculated using the missing data algorithm until the actual metrology data becomes available. Another advantage of the PLS model is that it can be used with the batch control algorithm outlined by Flores-Cerrillo and MacGregor [23] to calculate control moves in the latent variable space. This allows the controller to take into account the correlations between CD measurements when making adjustments to the future processing steps.

## 5.1 PLS Modeling

A PLS model is obtained by taking data matrices $X$ and $Y$ and projecting them onto lower dimensional subspaces. The subspaces used to represent the original data are known as latent variables. The input and output data matrices can then be represented in latent variable form as shown in Equation 5.1. A detailed description of the NIPALS algorithm used for PLS model building can be found in [27].

$$X = TP^T + E \tag{5.1}$$

$$Y = TQ^T + F$$

Here $T = XR$ is the latent variable score matrix that captures most of the variance, $P$ and $Q$ are the loading matrices, and $E$ and $F$ are the residuals. The number of latent variables used in the PLS model is determined by the user. There are many methods of selecting the optimum number of latent variables [40, 69, 96] and the cross validation with predicted error sum of squares (PRESS) [96] was used for the models in this chapter.

## 5.2 Missing Data Algorithm

Data availability is instrumental in making good control decisions and control moves. The EPC algorithm uses data from complete processing steps and expected values for subsequent processing steps in making control decisions. Metrology delays and the fact that not all wafers are measured from every lot can prevent data from being available for model building, as well as when it is needed at the next processing step for control move decisions. The PLS model provides a convenient format for model building with missing data, estimating the missing metrology data, and estimating the expected values for future processing steps by using the correlation structure between variables contained in the model.

### 5.2.1 Missing Data Replacement with an Existing Model

The missing data algorithm estimates the missing data by determining values for the missing variables that minimize the residuals of the reconstructed variables projected onto the model. To do this a matrix is formed where each of the missing variables is represented by a unit column vector in the direction corresponding to the missing variable. These vectors are combined to form matrix $E_i$, that represents the directions of all the missing variables. The available variables are contained in a column vector, $x_a$, that has zeros in place of the missing variables, and the missing variables are in vector, $z_m$. The reconstructed data set, $x^r$, is represented by Equation 5.2.

$$x^r = x_a + E_i z_m \tag{5.2}$$

The residuals of the reconstructed variables projected onto the model, $\tilde{x}^r$, are shown in Equation 5.3.

$$\tilde{x}^r = \tilde{x_a} + \tilde{E}_i z_m \tag{5.3}$$

where

$$\tilde{x_a} = (I - PR^T)x_a \tag{5.4}$$

$$\tilde{E}_i = (I - PR^T)E_i \tag{5.5}$$

To reconstruct the missing variables the residuals are minimized. The solution to the minimization, with some manipulation, yields the reconstructed variable calculation given in Equation 5.6.

$$x^r = (I - E_i(\tilde{E_i}^T \tilde{E_i})^{-1} \tilde{E_i}^T)x_a \tag{5.6}$$

The usefulness of the reconstructed variables depends on the strength of the correlation between the missing and available variables. This can be evaluated by calculating the variance of reconstruction error (VRE) for the missing variable as described in [69] and comparing it to the variance of the historical data for the same variable. If the VRE is less than the historical data variance, then the estimates will yield better predictions than using the historical average, otherwise the historical average should be used.

### 5.2.2 Model Building with Missing Data

Most fabrication facilities measure CD data for a subset of wafers in each lot because it is too time consuming and costly to do otherwise and because the subset of wafers measured for each processing step is not the same. This is problematic when building models based on historical data because regression models require full matrices of data. The PLS modeling algorithm, NIPALS, can be modified with the single component projection method for missing data replacement to allow model building with missing data. The modifications to the NIPALS algorithm are described in [27, 61].

## 5.3 Latent Variable Control

The correlations between CD measurements are useful because they make it possible to use the missing data techniques outlined in Section 5.2.1, but they are problematic when determining control moves. If the correlations are not taken into account when calculating the optimal set points for the subsequent processing steps then changing the next step can cause unanticipated, and sub-optimal, changes in the later processing steps. In order to factor in the correlation in CD variables a method for calculating the optimal set points in the latent variable space is used [24]. This method was originally used to control batch product quality through trajectory manipulation but it can be easily integrated into the EPC algorithm. The PLS model scores, $\hat{t}^T$, are calculated using the data from completed steps, $x_1$, estimates or nominal set points for the later processing steps, $\hat{x_2}$, and the PLS model matrix $R$ as shown in Equation 5.7.

$$\hat{t}^T = [x_1 \quad \hat{x_2}]R \tag{5.7}$$

The optimal set points are calculated by minimizing the difference between the predicted and targeted electrical parameters and the minimization finds the solution in terms of score changes. This formulation is shown in Equation 5.8.

$$\begin{aligned} &\min_{\Delta t} (\hat{y} - y_{target})^T (\hat{y} - y_{target}) + \Delta t^T W \Delta t \\ &s.t. \quad \hat{y}^T = (\Delta t + \hat{t})^T Q^T \end{aligned} \tag{5.8}$$

Here $\Delta t^T = t^T - \hat{t}^T$ is the vector of optimal changes in the score, $t^T$ are the new optimal scores, $Q$ is a matrix from the PLS model, and $W$ is a weighting matrix for the movement suppression term of the objective function. Once the optimal score changes are computed they need to be mapped back to optimal changes in the set points for the subsequent processing steps. This is accomplished using Equation 5.9.

$$\hat{x_2}^T = (t^T_{optimal} - x^T_1 W_1)(P^T_2 W_2)^{-1} P^T_2 \tag{5.9}$$

Here $W_1$, $W_2$, and $P_2$ are created by partitioning the PLS model matrices, $W$ and $P$, to match the partitioned data matrices $x_1$ and $\hat{x_2}$.

## 5.4 PLS Model Updating

The PLS model needs to be updated in order for it to adequately reflect slow changes in the process. The model can be updated by recalculating a model bias term after a few lots have been processed. The model bias term is the weighted average of the previous model bias term and the difference between the predicted and actual electrical parameter values for the lots that completed processing since the last update. The PLS model with the bias update is shown in Equation 5.10 and the model bias term calculation is given in Equation 5.11.

$$\hat{Y_k} = X_k R Q^T + B_k \tag{5.10}$$

$$B_k = \lambda_{MU} B_{k-1} + (1 - \lambda_{MU}) \frac{1}{w} \sum_{i=1}^{w} (Y_i - \hat{Y_i}) \tag{5.11}$$

Here $B_k$ is the model bias term, $w$ is the number of lots that have completed processing since the last update, and $\lambda_{MU}$ is a weighting factor.

If the process has changed significantly and the entire model needs to be updated instead of just the bias term there are two approaches. One is to recalculate the PLS model with more current data and the other is to use the recursive partial least squares method described in [66]. The latter method requires less computing power since it does not involve inverting the entire data matrix.

## 5.5 Lot Level Control

The lack of CD data available for individual wafers in a lot, combined with the inability to make wafer level adjustments at each processing step, creates an environment where it is difficult to make control decisions and control moves at the wafer level. This problem can be solved by predicting electrical parameter values and calculating control moves for all of the wafers available in the lot and then averaging the values. The implemented control moves would be the average of the control moves for the individual wafers whose data is available in each lot. Similarly, the control move decisions will be based on the average of the predicted electrical parameter values for each wafer whose data is available in the lot. These calculations are shown in Equation 5.12 and Equation 5.13, where $\hat{y}$ represents the predicted electrical parameter values, $\hat{x_2}$ contains the optimal set points for the subsequent processing steps, and $n$ is the number of wafers in the current lot that have CD measurements available.

$$\hat{y}_{lot} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i \tag{5.12}$$

$$\hat{x}_{2,lot} = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_{2,i} \tag{5.13}$$

The industrial data shown in Figure 5.1 justifies using lot averages for control decisions and control moves. The wafer data, represented by the blue line, has been plotted so that wafers within the same lot are connected together. The red line represents the lot average for each of the lots and the black lines are the $\pm 1\sigma$ lines that are at 1 and -1 respectively, since the data has been normalized. By looking at the plot it is easy to see that the averages for some of the lots fall outside of the standard deviation lines. Moving the averages of these lots, via lot to lot control, will reduce the variability in the process.

## 5.6 Simulation

Two simulations are performed to determine how well the proposed enhancements to the EPC algorithm perform. The first simulation compares the performance of the EPC algorithm described in Chapter 4 with the same algorithm combined with missing value calculations for prediction purposes. It also compares those two implementations with the latent variable EPC algorithm combined with missing value calculations for predictions. The second simula-

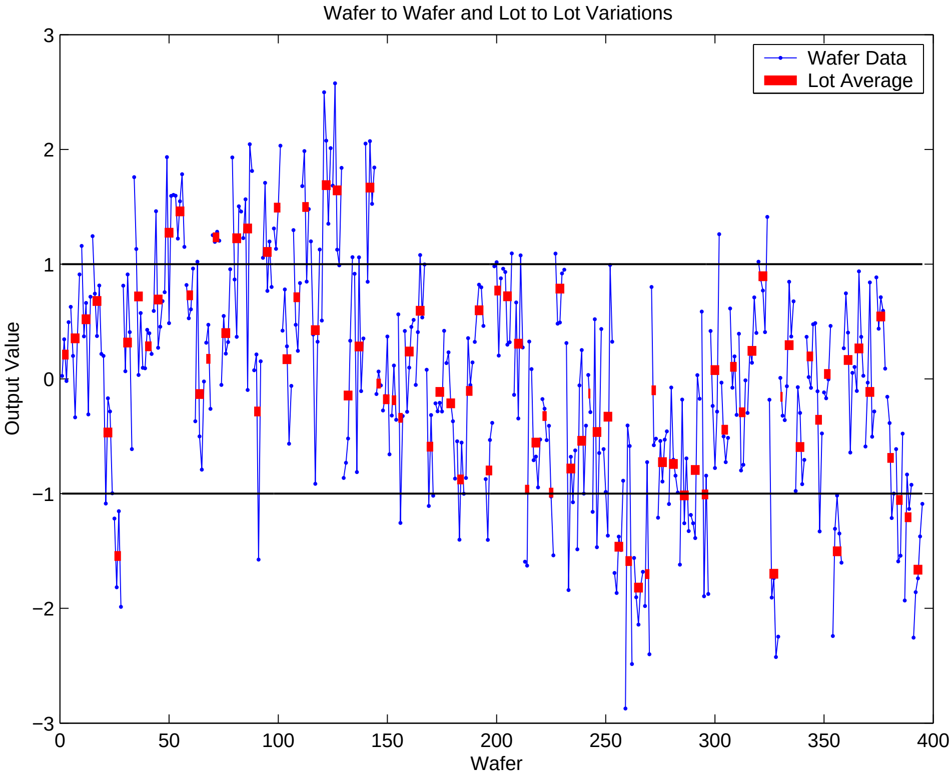


Figure 5.1: Industrial data showing wafer to wafer and lot to lot variations

tion investigates how well lot level control works with missing measurements and the latent variable EPC algorithm.

### 5.6.1 EPC Algorithm Comparison

The purpose of this simulation is to see how much improvement the EPC algorithm enhancements provide. The simulation has ten inputs and three outputs. The inputs are correlated and adjustments are made starting at the fourth processing step. For this simulation CD data is available for every wafer after each processing step and there is the ability to make wafer level adjustments. The simulation generates 1000 data points and the first 500 are used for model building. The last 500 points are generated with no control and then regenerated with each of the different controllers turned on so that the results can be compared directly. Test case one is the EPC algorithm described in Chapter 4. This algorithm formulation uses the least squares model for prediction and control. It also uses the nominal set point values for later processing steps when making predictions that are used to make control move decisions. Test case two is the same EPC algorithm as in test case one but instead of using nominal set points for predictions that facilitate control move decisions it uses the PLS missing data algorithm to calculate values for later processing steps. Test case three uses the latent variable algorithm described in Section 5.3 and uses the nominal set points for prediction and control move decisions. Test case four uses the latent variable control algorithm combined with the PLS missing data algorithm to calculate values for the later processing

Table 5.1: Test Case Descriptions

| Test Case | EPC Algorithm Method | Subsequent Processing Data |
|---|---|---|
| 1 | Least Squares | Nominal Set Point |
| 2 | Least Squares | PLS Missing Data |
| 3 | Latent Variable | Nominal Set Point |
| 4 | Latent Variable | PLS Missing Data |

Table 5.2: Results for the Latent Variable EPC Algorithm Comparison Simulation

| Test Case | | y1 STD | y2 STD | y3 STD |
|---|---|---|---|---|
| No Control | Data | 2.1360 | 2.0411 | 2.3136 |
| TC 1 | Data | 1.8548 | 1.7667 | 2.0058 |
| | % Improvement | 13.2% | 13.4% | 13.3% |
| TC 2 | Data | 1.7282 | 1.6396 | 1.8916 |
| | % Improvement | 19.1% | 19.7% | 18.2% |
| TC 3 | Data | 1.7793 | 1.7104 | 2.0036 |
| | % Improvement | 16.7% | 16.2% | 13.4% |
| TC 4 | Data | 1.6115 | 1.4969 | 1.7332 |
| | % Improvement | 24.6% | 26.7% | 25.1% |

steps. In all of the test cases involving the missing data algorithm the missing data estimates are only used when the VRE is less than the historical variation for the parameter. If the VRE is greater than the historical variation then the nominal set point is used instead. The four test cases are summarized in Table 5.1 and the results are summarized in Table 5.2.

The simulation results show that using the missing data algorithm yields significant improvements for both the least squares and the latent variable implementation of the EPC algorithm. This makes sense because the algorithm bases its control decisions off of the electrical parameter predictions and the better the predictions are the more likely control move decisions are the appropriate ones. The simulation results also show that the latent variable EPC algorithm combined with the PLS missing data algorithm for predictions yields the most improvements in environments where the CD measurements are correlated.

### 5.6.2 Latent EPC with Lot to Lot Adjustment

The purpose of this simulation is to evaluate lot to lot adjustments versus wafer to wafer adjustments when only a subset of the wafers in a lot have measurement data available. This simulation uses the latent variable EPC algorithm and the missing data prediction algorithm as well.

The manufacturing process that is simulated has 13 processing steps and two electrical parameter outputs to control. Three of the processing steps are independent and they have lower level R2R controllers acting at those steps. The outputs for those steps are calculated in the same manner as the R2R control simulation described in Section 4.4.1. The other 10 processing steps are correlated with the three independent processing steps and do not have R2R controllers. Instead, they allow for manual adjustment of the set point. The outputs for those steps are calculated using Equation 5.14.

$$x_{out} = CX_{ind} + x_{sp} + \varepsilon_x \tag{5.14}$$

Here $x_{out}$ is the output for the processing step, $X_{ind}$ is a matrix of the three independent processing step values, $C$ is the gain matrix relating the independent processing steps to the dependent ones, $x_{sp}$ is the set point for the processing step, and $\varepsilon$ is normally distributed process noise. The electrical parameter outputs are calculated the same way as in the fab-wide simulation described in Section 4.4.2 except that another noise term is added. The noise term is the same for each lot of wafers and is used to model the lot by lot variation seen in the industrial data. The electrical parameter calculation with the added noise term is shown in Equation 5.15

$$Y_{ep,ij} = \theta X_{cd,ij} + d_{ij} + m_{error,ij}\varepsilon_{ep} + l_{error,j}\varepsilon_{l} \quad i = 1...24 \quad j = 1...m. \quad (5.15)$$

Here $l_{error,j} = 0.2$ is the standard deviation of the normally distributed lot error and $m$ is the number of lots. The other variables and their values are defined in Section 4.4.2.

To simulate the environment where only a subset of the wafers from each lot are measured and available for use, the data for the first five wafers of every lot are used for model building as well as control decisions and control move calculations. To model the process, data is generated with no controller and a PLS model is built from that data. Adjustments to the process can be made starting at the fifth processing step and at all of the subsequent processing steps. The wafer to wafer noise and lot to noise levels are similar to the noise levels shown in the industrial data. The average of the predicted

Table 5.3: Test Case Descriptions

| Control Type | | Output 1 | Output 2 |
|---|---|---|---|
| Wafer to Wafer | Without Control | 0.6258 | 0.6181 |
| | With Control | 0.4530 | 0.4498 |
| | % Improvement | 27.6% | 27.2% |
| Lot to Lot | Without Control | 0.6292 | 0.6479 |
| | With Control | 0.4995 | 0.5058 |
| | % Improvement | 20.6% | 21.2% |

electrical parameters for the five wafers in each lot are used to determine if control moves are made. If a control move is made the average of the control move calculated for each of the five wafers is applied to the entire lot. In Table 5.3 the performance of the simulation is compared to data for the same process but with wafer level adjustments and plots of the output for wafer level control and lot level control are shown in Figure 5.2 and Figure 5.3 respectively.

The results of the simulation show that wafer to wafer control yields better results that lot to lot control. This is expected since wafer to wafer control can fix problematic wafers within a lot and not just entire lots that are problematic. The simulation also shows that lot to lot control can accomplish significant improvements and is a viable strategy for controlling manufacturing processes when wafer to wafer control is not an option.

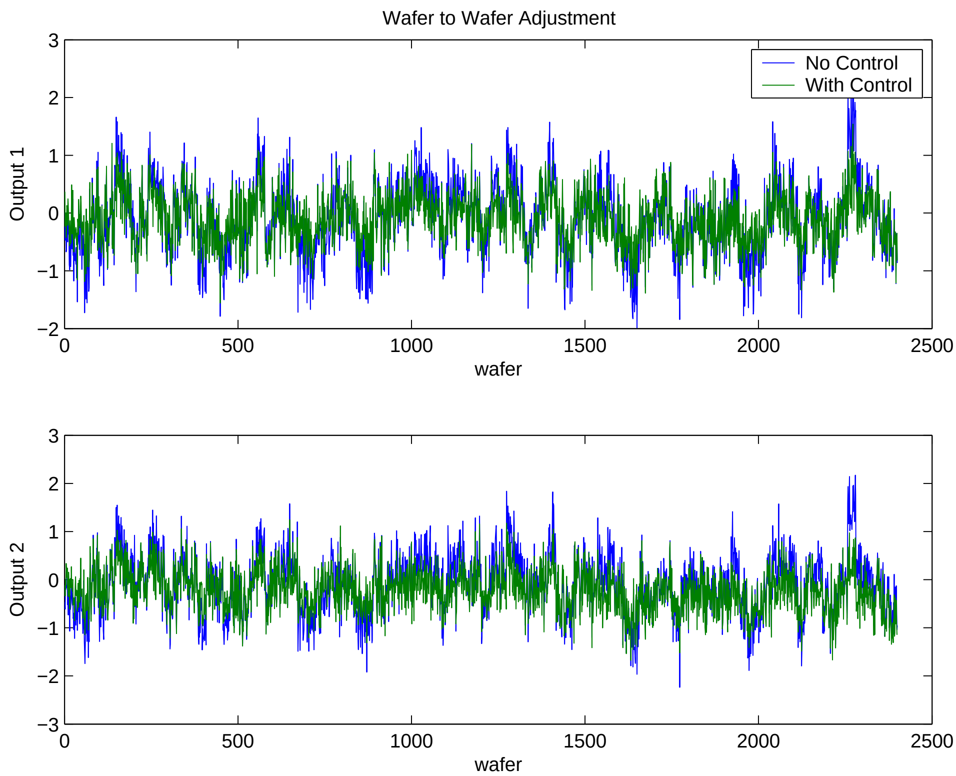


Figure 5.2: Output data with wafer to wafer EPC control applied

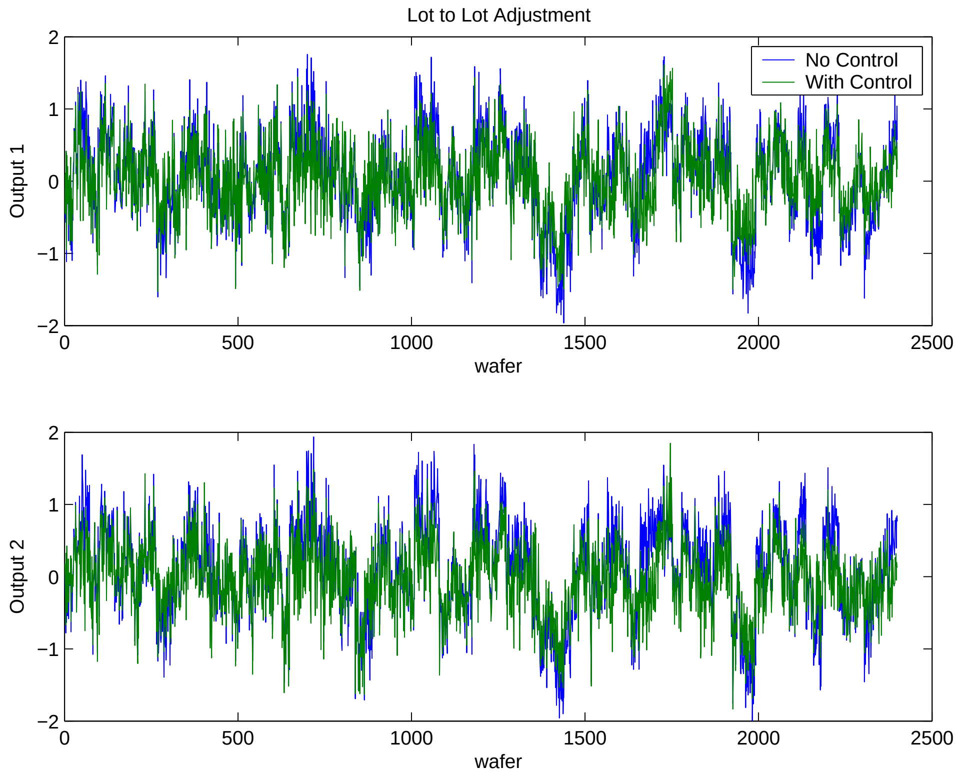


Figure 5.3: Output data with lot to lot EPC control applied

## 5.7 Limitations

The major limitation of the latent variable EPC algorithm has to do with invertibility issues in the calculation of the optimal set points for the subsequent processing steps. When the number of missing variables to estimate is less than the number of latent variables in the PLS model it is possible for the $P_2^T R_2$ matrix to be rank deficient. The inverse of this term is used in the optimal set point calculation given in Equation 5.9. There are a couple of solutions to this problem. The first is simply to not make adjustments at the remaining steps. This option is not ideal because these are the last few chances to make adjustments. The second option is to collect data for processing steps that occur after that last step where adjustments should be made. If these later steps are correlated with earlier ones they can be put into the model and used make the number of steps remaining greater than the number of latent variables so that inversion is not a problem.

# Chapter 6

# Other EPC Implementations

Two alternative methods to implementing the EPC algorithm are discussed in this chapter. The first method is a technique used for multi-step control in photolithographic sequences that is applied to fab-wide control. The second method is a Bayesian approach to the EPC algorithm that addresses the stochastic nature of semiconductor manufacturing processes.

## 6.1 Cost Function EPC Implementation

The cost function approach to multi-step control was first proposed by Leang et al. [47] to control a sequence of photolithographic processing steps. Empirical models relating the inputs and outputs of each processing step were used to perform a Monte Carlo simulation. The results of the simulation yielded acceptable input regions corresponding to acceptable output regions for each of the processing steps. The limits of the acceptable regions are used to determine the weights, or costs, used in the objective function that is set up for calculating optimal input values for the next processing step. For independent process outputs the cost function is defined in Equation 6.1.

$$Cost = \sum_{i=1}^{p} k_i (y_i - y_{i,target})^2 \tag{6.1}$$

Here $y$ represents the process outputs, $k$ are the scaling coefficients, and $p$ is the number of process outputs. The scaling coefficients are calculated so that the cost function equals one when a process output is at the edge of its acceptable region, assuming the other process outputs are on target. If the outputs are correlated, as is often the case, principle component analysis (PCA) is used to decorrelate the outputs. PCA is similar to PLS because the algorithm projects the data onto lower dimensional subspaces. These subspaces are orthogonal so that the new variables that represent the data are independent. PCA is discussed in detail in [98]. When the outputs are correlated the cost function expression is altered to reflect the transformation and is given in Equation 6.2.

$$Cost = \sum_{i=1}^{p} \sum_{j=1}^{p} k_{ij} (y_i - y_{i,target})(y_j - y_{j,target}) \tag{6.2}$$

Here the scaling coefficients, $k_{ij}$, are calculated in the same manner as the $k_i$ coefficients in Equation 6.1 except the PCA transformation is involved. In order to calculate the optimal set points for the later processing steps the cost function is minimized over the remaining inputs. To facilitate this the empirical equation relating the inputs and the outputs, $y_i = f(x_1, ..., x_n)$, is substituted into the cost function for $y_i$ using the available $x$ data from completed processing steps.

### 6.1.1 Cost Function Simulation

The purpose of this simulation is to evaluate the effectiveness of the cost function objective compared to the other EPC formulations already presented. In order to do this the same manufacturing simulation described in Section 5.6.1 is used. This simulation has ten manufacturing steps and three electrical parameter outputs to control and the EPC algorithm is performed on a wafer to wafer basis. Since the outputs of this simulation are correlated a PCA model of the outputs is built, $Y = TP^T$, and the cost function in the latent variable space is defined in Equation 6.3.

$$Cost = \sum_{m=1}^{n} k_{m,pca}(yp_m - y_{target}p_m)^2 \tag{6.3}$$

Here $k_{pca}$ are the scaling coefficients for the latent variable cost function, $p_m$ is the $m^{th}$ column of the loading matrix, $P$, from the PCA model, $y_{target}$ are the target values for the outputs, and $n$ is the number of latent variables. The scaling coefficients are calculated using the cost function by setting the cost to one when an output $m$ is at its upper limit, $y_{m,ul}$, and all the other outputs are on target. This calculation is shown in Equation 6.4.

$$k_{m,pca} = \frac{1}{(y_{m,ul}p_m - y_{m,target}p_m)^2} \quad m = 1...n \tag{6.4}$$

The scaling coefficients are transformed into the decorrelated scaling coefficients, $k_{ij}$, according to Equation 6.5.

$$k_{ij} = \sum_{m=1}^{n} k_{m,pca} p_{mi} p_{mj} \tag{6.5}$$

Here $p_{ij}$ is the $i^{th}j^{th}$ entry in the loading matrix, $P$. Once the decorrelated scaling coefficients are calculated the simulation is performed using the cost function defined in Equation 6.2. For this simulation the target values for the outputs were the historical averages and the upper limits for the outputs were one standard deviation from the historical averages.

The test cases investigated in the simulation are summarized in Table 6.1. Test case one uses the least squares implementation with the PLS missing data algorithm for prediction, test case two uses the latent EPC algorithm with the PLS missing data algorithm for prediction, test case three uses the cost function method with three latent variables and the PLS missing data algorithm, and test case four uses the cost function method with two latent variables and the PLS missing data algorithm. The results are summarized in Table 6.2.

Table 6.1: Test Case Descriptions for Cost Function Comparison

| Test Case | EPC Algorithm Method | Subsequent Processing Data |
|---|---|---|
| 1 | Least Squares | PLS Missing Data |
| 2 | Latent Variable | PLS Missing Data |
| 3 | Cost Function 3 LV | PLS Missing Data |
| 4 | Cost Function 2 LV | PLS Missing Data |

Table 6.2: Results for the Cost Function Comparison Simulation

| Test Case | | y1 STD | y2 STD | y3 STD |
|---|---|---|---|---|
| No Control | Data | 2.1360 | 2.0411 | 2.3136 |
| TC 1 | Data | 1.7282 | 1.6396 | 1.8916 |
| | % Improvement | 19.1% | 19.7% | 18.2% |
| TC 2 | Data | 1.6115 | 1.4969 | 1.7332 |
| | % Improvement | 24.6% | 26.7% | 25.1% |
| TC 3 | Data | 1.3685 | 1.3072 | 2.2128 |
| | % Improvement | 36.0% | 36.0% | 4.38% |
| TC 4 | Data | 1.1757 | 1.4105 | 1.5262 |
| | % Improvement | 45.0% | 30.9% | 34.0% |

The results show that the cost function approach has the ability to outperform the latent EPC algorithm in all three outputs but its success is dependent on the number of latent variables chosen for the PCA model of the outputs. In this particular simulation only two independent directions of variance were needed to represent the three outputs. When three latent variables were used, the third latent variable only modeled the noise level in the system. Using the loading vector for the third latent variable to determine weights for the system was not meaningful and the result is poor improvement in the third output. The performance increases for the third and first variable and decreases a little for the second variable when the two meaningful latent variables are used for weighting calculations.

## 6.2 Bayesian EPC Implementation

The Bayesian approach to the EPC algorithm addresses the stochastic nature of the semiconductor manufacturing environment by dealing with the probabilities of events occurring instead of calculating deterministic outcomes for each event. In this formulation the CD measurements for each processing step are considered random variables. The electrical parameter outputs are functions of the CD measurement random variables, and are themselves random variables. The functional relationship between the CD random variables and the electrical parameter random variables can be written as shown in Equation 6.6.

$$Y = g(X_1, ..., X_n) \tag{6.6}$$

Here, $Y$, is the electrical parameter random variable, $g$ is the function relating the CD measurement random variables to the electrical parameter random variables, $X$ are the CD measurement random variables, and $n$ is the number of processing steps. Assuming the joint probability density function (PDF) for the CD measurement random variables is known the electrical parameter PDF can be calculated [63]. The standard calculation can be cumbersome to work with and an alternative method based on the Dirac delta function is proposed by Au and Tam [1] and shown in Equation 6.7. This functional form can be easily extended to joint distributions of several functions as described in [41], which is a useful property for dealing with more than one electrical parameter output.

$$f_Y(y) = \int_{\mathbb{R}^n} f_X(\mathbf{x})\delta[y - g(\mathbf{x})]d\mathbf{x} \tag{6.7}$$

Here $f_Y(y)$ represents the electrical parameter probability density function, $\mathbf{x}$ contains the vector of input variables, $f_X(\mathbf{x})$ represents the joint probability density function for the inputs, and $\mathbb{R}^n$ shows that this is an $n$ dimensional integral over all of the inputs.

Adapting the EPC algorithm for the Bayesian approach involves using probabilities to make control moves and control move decisions. Instead of using a deterministic model and data from completed processing steps to predict the electrical parameters values, Equation 6.7 is modified to calculate the probability that the electrical parameters will be within a specific range, given the CD data from completed processing steps. This range can be based on product specifications or other metrics. The electrical parameter PDF, given data from completed steps, is calculated as shown in Equation 6.8.

$$f_Y(y|x_1, ..., x_i) = \frac{\int_{\mathbb{R}^n} f_X(\mathbf{x})\delta[y - g(\mathbf{x})]dx_{i+1}...dx_n}{\int_{\mathbb{R}^n} f_X(\mathbf{x})dx_{i+1}...dx_n} \tag{6.8}$$

Here $x_1, ..., x_i$ is the available data from the $i$ completed processing steps and the integrals are taken over the remaining $i + 1$ to $n$ subsequent processing steps. The conditional PDF for the electrical parameters, given available processing data, is integrated over the desired range, $y_{ll}$ to $y_{ul}$, to determine the probability that the electrical parameters will be within that range, as shown in Equation 6.9.

$$P\{y_{ll} < y < y_{ul}\} = \int_{y_{ll}}^{y_{ul}} f_Y(y|x_1, ..., x_i)dy \tag{6.9}$$

Based off of the calculated probability, the decision to make a control move can be made. If the CD measurements from completed processing steps have known measurement errors associated with them this can be factored in to the PDF for the electrical parameters by performing all $n$ integrals and integrating between the error bounds for each of the completed steps instead of evaluating the PDF at the data points.

If a control action is necessary the control move can also be determined by using the electrical parameter conditional PDF given in Equation 6.8. This formulation is shown in Equation 6.10. For this purpose Equation 6.8 is evaluated with the conditions of the known CD data, $x_1 ... x_i$, and the value for the next processing step, $x_{i+1}$. Since $x_{i+1}$ is not known the electrical parameter PDF is a function of $x_{i+1}$ as well as $y$. Then the probability of the electrical parameters being in the range $y_{ll}$ to $y_{ul}$ is determined using Equation 6.9 and the resulting probability will be a function of the unknown processing step value, $x_{i+1}$. This function can be maximized over $x_{i+1}$ to determine the optimal set point for the next processing step.

$$f_Y(y|x_1, ..., x_{i+1}) = f(y, x_{i+1}) \tag{6.10}$$
$$\max_{x_{i+1}} P\{y_{ll} < y < y_{ul}\} = f(x_{i+1}) = \int_{y_{ll}}^{y_{ul}} f_Y(y|x_1, ..., x_{i+1})dy$$

The process described here only calculates the optimal move for the next step. If it is advantageous to distribute the corrections over multiple steps then the conditional electrical parameter probability should be determined with all of the steps over which adjustments need to be made included in the conditional statement. If all of the process inputs are included in the determination of the set point for the subsequent processing step then the maximization can take place using the functional relationship, $Y = g(X_1, ..., X_n)$, instead.

### 6.2.1 Bayesian EPC Simulation

The purpose of this simulation is to determine the viability of the Bayesian formulation of the EPC algorithm. The results of this simulation are not directly comparable to the simulations done in Section 5.6.1 and Section 6.1.1 because assumptions are made to simplify the process that is being simulated. The assumptions are necessary because of the complexity of the integrals involved in the Bayesian formulation. The system under simulation has four processing steps and one electrical parameter output to control and adjustments are made at the third and fourth processing steps. The CD measurements from each of the four processing steps are assumed to be independent, normally distributed, random variables with different means and variances. The independence assumption simplifies the computation of the electrical parameter probabilities and the optimal set point calculations. The electrical parameter random variable is related to the CD measurement random variables by the function shown in Equation 6.11.

$$Y = g(X_1, ..., X_n) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 \qquad (6.11)$$

Since the processing steps are independent, the electrical parameter conditional PDF calculation is simplified because the joint distribution, $f_X(\mathbf{x})$, is equal to the product of the individual PDFs for the each of the processing steps. Using the independent normal distributions, the resulting electrical parameter conditional PDF, given data for steps one and two, is shown in Equation 6.12 and the electrical parameter PDF, given data from steps one, two, and three, is shown in Equation 6.13.

$$\begin{aligned} f_Y(y|x_1,x_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_3}(x_3)f_{X_4}(x_4)\delta[(y-\alpha\mathbf{x})]dx_3dx_4 \qquad (6.12) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_3}(x_3)f_{X_4}(x_4)\delta[y-\alpha_1x_1-\alpha_2x_2-\alpha_3x_3-\alpha_4x_4]dx_3dx_4 \end{aligned}$$

$$\begin{aligned} f_Y(y|x_1,x_2,x_3) &= \int_{-\infty}^{\infty} f_{X_4}(x_4)\delta[(y-\alpha\mathbf{x})]dx_4 \qquad (6.13) \\ &= \int_{-\infty}^{\infty} f_{X_4}(x_4)\delta[y-\alpha_1x_1-\alpha_2x_2-\alpha_3x_3-\alpha_4x_4]dx_4 \end{aligned}$$

$$f_K(x_k) = \frac{1}{\sqrt{2\pi}\sigma_k}e^{-\frac{(x_k-\mu_k)^2}{2\sigma_k^2}}$$

The first integral is performed by recalling that the integral of the product of a function and a delta function is given by Equation 6.14 [31].

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)dx = f(a) \qquad (6.14)$$

Using this property of the delta function and substituting the normal distributions the integrals are evaluated. The PDFs are now given in Equation 6.15 and Equation 6.16.

$$\begin{aligned} f_Y(y|x_1,x_2) &= \int_{-\infty}^{\infty} f_{X_3}\left(\frac{y-\alpha_1x_1-\alpha_2x_2-\alpha_4x_4}{\alpha_3}\right) f_{X_4}(x_4)dx_4 \\ &= \frac{1}{2\pi\sigma_3\sigma_4}\int_{-\infty}^{\infty} e^{-\frac{\left(\frac{y-\alpha_1x_1-\alpha_2x_2-\alpha_4x_4}{\alpha_3}-\mu_3\right)^2}{2\sigma_3^2}-\frac{(x_4-\mu_4)^2}{2\sigma_4^2}} dx_4 \end{aligned} \tag{6.15}$$

$$\begin{aligned} f_Y(y|x_1,x_2,x_3) &= f_{X_4}\left(\frac{y-\alpha_1x_1-\alpha_2x_2-\alpha_3x_3}{\alpha_4}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_4\alpha_4} e^{-\frac{(y-\alpha_1x_1-\alpha_2x_2-\alpha_3x_3-\mu_4\alpha_4)^2}{2\sigma_4^2\alpha_4^2}} \end{aligned} \tag{6.16}$$

The calculation of the $P_Y(y|x_1,x_2,x_3)$ is complete but the PDF for the $P_Y(y|x_1,x_2)$ requires one more integral calculation. The exponent is a quadratic function in $x_4$ and the integral of an exponential with a quadratic exponent is given in Equation 6.17 [85].

$$\int_{-\infty}^{\infty} e^{-(ax^2+bx+c)}dx = \sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}-c} \tag{6.17}$$

Rearranging the exponent, using Equation 6.17, and performing tedious algebra results in the electrical parameter PDF, given data from steps one and two, as shown in Equation 6.18.

$$f_Y(y|x_1,x_2) = \frac{\alpha_3}{\sqrt{2\pi(\sigma_3^2\alpha_3^2+\sigma_4^2\alpha_4^2)}} e^{-\frac{(y-\alpha_1 x_1-\alpha_2 x_2-\mu_3\alpha_3)^2+\mu_4^2\alpha_4^2}{2(\sigma_3^2\alpha_3^2+\sigma_4^2\alpha_4^2)}} \tag{6.18}$$

Equation 6.18 and Equation 6.16 are integrated between the upper and lower limits of the electrical parameter output to determine the probability that the electrical parameter is within the desired range, given the available data. In this simulation the upper and lower limits for the electrical parameter are one standard deviation above and below the electrical parameter process mean of the simulation with no control. For this simulation, the control move threshold after each step is different, depending on the available data. To determine what the threshold should be the probability of the electrical parameters being on target was calculated hundreds of times and the distributions were examined. Based on the distributions a threshold was selected. When step one and step two data is available the probability that the electrical parameter is on target needs to be less than 0.4 before a control move is made. When data for steps one, two, and three are available the probability needs to be less than 0.65 before a control move is made.

If a control move needs to be made, the optimal set point for the next step is determined by finding the value that maximizes the probability that the electrical parameter will be within its specified range, conditional on the available data, as shown in Equation 6.10. For this simulation the optimal set point for step three is the value of $x_3$ that is the solution to Equation 6.19.

Table 6.3: Bayesian EPC Performance

| | Mean | STD |
|---|---|---|
| Without Control | 0.4545 | 0.4988 |
| With Control | 0.4862 | 0.2645 |
| % Improvement | | 47.0% |

$$\max_{x_3} f_Y(y|x_1,x_2,x_3) = \frac{1}{\sqrt{2\pi}\sigma_4\alpha_4} e^{-\frac{(y-\alpha_1x_1-\alpha_2x_2-\alpha_3x_3-\mu_4\alpha_4)^2}{2\sigma_4^2\alpha_4^2}} \tag{6.19}$$

The optimal set point for step four is calculated according to Equation 6.20

$$x_{4,sp} = \frac{y_{target} - \alpha_1x_1 - \alpha_2x_2 - \alpha_3x_3}{\alpha_4} \tag{6.20}$$

The results of this simulation are shown in Table 6.3. The simulation results indicate that the standard deviation of the electrical parameter output can be reduced by up to 47%, which is a significant improvement for any manufacturing process. Given this simplified and ideal scenario, it is unrealistic to expect such dramatic results if this method were applied in a real manufacturing environment.

### 6.2.2 Limitations

The limitations of the Bayesian EPC implementation have to do with the ability to determine the PDFs associated with the process and the complexity of the integrals necessary to execute this algorithm. The independence assumption in the simulation simplifies the calculations but is not a practical assumption. Developing probability distributions requires historical data and the higher the dimensionality of the process the more data is necessary. Despite multiple methods for density estimation [80], determining the joint PDF for processes with only a few correlated manufacturing steps could prove difficult. Assuming the density functions are available implementation could still be challenging. If the density estimates can be integrated numerically it is feasible to use this method but if the PDFs are functions, as in the simulation example here, the complexity of the integrals that need to be performed could prohibit the implementation of this algorithm.

# Chapter 7

# Conclusion

This dissertation has presented detailed implementation strategies for a fab-wide electrical parameter controller, including modeling techniques, optimal set point calculations, and model updating methods for each one. Each of the EPC algorithms is evaluated and its effectiveness is compared to that of the other algorithms, and their limitations are discussed.

The least squares EPC algorithm uses a least squares model for prediction and control. Control moves are determined by minimizing the difference between predicted and targeted electrical parameters. The least squares EPC algorithm was shown to effectively reduce electrical parameter variations through simulations, as well as an industrial case study. A model quality study was also preformed and showed that models with $R^2$ values as low as 0.6 are adequate for implementing the least squares EPC algorithm. The least squares EPC algorithm achieves its best results when the inputs are not significantly correlated.

In most manufacturing environments the processing steps are correlated and in these situations the latent variable EPC method yields better control results. The latent variable EPC algorithm calculates control moves in the latent variable space so that the variable correlations are taken into account. The latent variable EPC method can be combined with missing data estimation to make better electrical parameter predictions. A simulation comparing the least squares EPC algorithm with and without using missing data estimation to the latent variable EPC algorithm with missing data estimation is performed. The results show that processes with correlated manufacturing steps are best controlled by the latent variable EPC algorithm with missing data estimation. A simulation is also used to examine the use of lot level control based off the measurements of a subset of wafers in each lot and the results show that lot level control is feasible.

A cost function based approach is presented that utilizes a user defined acceptable input region and the cost associated with each input is used to re-center the process when control adjustments need to be made. When there are correlated outputs a PCA transformation is applied to the outputs so that each one is independent before determining the costs. A simulation is performed and the cost function formulation can achieve significant reductions in electrical parameter variance compared to the latent variable EPC algorithm when the correct number of variables are chosen for the PCA model.

A Bayesian EPC algorithm is described that bases control moves and control move decisions on probabilities instead of deterministic predictions. The probability that the electrical parameters are within a user defined acceptable range is determined, conditional on the data from the completed manufacturing steps. This probability is maximized to determine the optimal control move for the next processing step. A simulation is performed and the results indicate that this method can effectively reduce electrical parameter variations. The practicality of the Bayesian EPC algorithm is severely limited due to the difficulty in obtaining probability density functions and the complex computations involved in the algorithms implementation.

This dissertation has presented detailed algorithmic implementations for fab-wide control and there is room for more work in the area. So far only empirical models have been used but the potential for physics based first principle models needs to be examined. This work does not outline how to integrate the fab-wide EPC controller with module level feedforward controllers that already combine multiple steps in manufacturing and this should be investigated. Lastly these ideas need to be implemented on a pilot manufacturing line to assess performance in a real manufacturing environment. Once this has been accomplished the algorithm can be refined further.

In conclusion, four methods for implementing fab-wide EPC have been proposed. Simulations and an industrial case study have been performed to show that each method has been successful at reducing electrical parameter variations, though some are significantly better than others depending on the nature of the process that is under control. The area of fab-wide EPC has the potential to greatly improve semiconductor manufacturing performance through increased yield and reduced cost and continued work in the area is needed to realize this potential.

# Bibliography


[1] C. Au and J. Tam. Transforming variables using the Dirac generalized function. *American Statistician*, 53(3):270–272, 1999.

[2] T. A. Badgwell, T. Breedik, S. G. Bushman, S. W. Butler, S. Chatterjee, T. F. Edgar, A. J. Toprac, and I. Trachtenberg. Modeling and control of microelectronics materials processing. *Computers and Chemical Engineering*, 19(1):1–41, 1995.

[3] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forcasting and Control, 3rd Ed.* Englewood Cliffs, NJ Prentice Hall, 1994.

[4] S. Bushman and S. Farrer. Scatterometry measurements for poly-gate etch process monitoring. In *Proceedings of Sematech AEC/APC Symposium*, 1997.

[5] S. W. Butler and J. Stefani. Supervisory run-to-run control of polysilicon gate etching using in-situ ellipsometry. *IEEE Transactions on Semiconductor Manufacturing*, 7:193–201, 1994.

[6] W. Campbell. *Model Predictive Run-to-run control of Chemical Mechanical Planerization.* PhD thesis, The University of Texas at Austin, 1999.

[7] J. P. Card, D. L. Sniderman, and C. Klimasauskas. Dynamic neural control for plasma etch processes. *IEEE Tasnactions on Neural Networks*, 8(4):883–901, 1997.

[8] N. Chaudhry, J. Moyne, and E. Rundensteiner. Active controller: Utilizing active databases for implementing multi-step control of semiconductor manufacturing. *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, C, 1998.

[9] A. Chen and R. S. Guo. Age-based double EWMA controller and its application to CMP processes. *IEEE Transactions on Semiconductor Manufacturing*, 14(1):11–19, 2001.

[10] P. Congdon. *Bayesian Statistical Modelling, 2nd Ed.* John Wiley & Sons, 2006.

[11] T. J. Crowley and K. Y. Choi. Experimental studies on optimal molecular weight control in a batch-free radical polymerization process. *Chemical Engineering Science*, 53(15):2769–2790, 1998.

[12] E. Del Castillo. Long run and transient analysis of a double EWMA feedback controller. *IIE Transactions*, 31:1157–1169, 1999.

[13] E. Del Castillo. *Statistical Process Adjustment for Quality Control.* New York, Wiley, 2002.

[14] E. Del Castillo and A. M. Hurwitz. Run-to-run process control: literature review and extensions. *Journal of Quality Technology*, 29:184–196, 1997.

[15] E. Del Castillo and J. Yeh. An adaptive run-to-run optimizing controller for linear and nonlinear semiconductor processes. *IEEE Transactions on Semiconductor Manufacturing*, 11-2:285–295, 1998.

[16] G. Dreyfus. *Neural Networks: Methodology and Applications*. Springer, 2005.

[17] T. F. Edgar, S. W. Butler, W. J. Campbell, C. Pfeiffer, C. Bode, S. B. Hwang, K. S. Balakrishnan, and J. Hahn. Automatic control in micorelectronics manufacturing: Practices, challenges, and possibilities. *Automatica*, 36:1567–1603, 2000.

[18] T. F. Edgar, W. J. Campbell, and C. Bode. Model-based control in microelectronics manufacturing. In *Conference on Decision and Control*, 1999.

[19] C. El Chemali, J. Freudenberg, M. Hankerson, and J. J. Bendik. Run-to-run critical dimension and sidewall angle lithography using the PROLITH simulator. *IEEE Transactions on Semiconductor Manufacturing*, 17(3):388–401, 2004.

[20] C. El Chemali, J. Freudenberg, M. Hankinson, W. Collison, and T. Ni. Critical dimension control of a plasma etch process by integrating feed-

forward and feedback run-to-run control. *Journal of Vacuum Science and Technology B*, 21(6):2304–2312, 2003.

[21] J. S. Fenner, M. K. Jeong, and J. C. Lu. Optimal automatic control of multistage production processes. *IEEE Transactions on Semiconductor Manufacturing*, 18(1):94–103, 2005.

[22] J. Flores-Cerrillo and J. F. MacGregor. Within-batch and batch-to-batch inferential-adaptive control of batch reactors: A PLS approach. *Industrial and Engineering Chemistry Research*, 42:3334–3345, 2003.

[23] J. Flores-Cerrillo and J. F. MacGregor. Control of batch product quality by trajectory manipulation using latent variable methods. *Journal of Process Control*, 14:539–553, 2004.

[24] J. Flores-Cerrillo and J. F. MacGregor. Control of batch product quality by trajectory manipulation using latent variable models. *Journal of Process Control*, 14:539–553, 2004.

[25] J. Flores-Cerrillo and J. F. MacGregor. Latent variable MPC for trajectory tracking in batch processes. *Journal of Process Control*, 15:651–663, 2005.

[26] J. F. Forbes, T. E. Marlin, and J. F. MacGregor. Model adequacy requirements for optimizing plant operations. *Computers and Chemical Engineering*, 18(6):497–510, 1994.

[27] P. Geladi and B. R. Kowalski. Partial least squares regression: A tutorial. *Analytica Chimica Acta*, 185:1–17, 1986.

[28] J.K. Ghosh. *An Introduction to Bayesian Analysis: Theory and Methods*. New York, Springer, 2006.

[29] R. Good. The control of electrical properties of semiconductors. In *Texas-Wisconsin Modeling and Control Consortium Spring Conference*, 2003.

[30] R. Good and S. J. Qin. On the stability of MIMO EWMA run-to-run controllers with metrology delay. *IEEE Transactions on Semiconductor Manufacturing*, 19(1):78–86, 2006.

[31] M. Greenburg. *Advanced Engineering Mathematics, 2nd Edition*. Prentice Hall, 1998.

[32] R. Guldi. Process optimization tweaking tool (POTT) and its application in controlling oxidation thickness. *IEEE Transactions on Semiconductor Manufacturing*, 2(2):54–59, 1989.

[33] M. Hankinson, T. Vincent, K. B. Irani, and P. P. Khargonekar. Integrated real-time and run-to-run control of etch depth in reactive ion etching. *IEEE Transactions on Semiconductor Manufacturing*, 10(1):121–130, 1997.

[34] C. A. Harrison, R. Good, D. Kadosh, and S. J. Qin. Multi-step supervisory control of flash memory device production via a simple first-principles model. In *AEC/APC Symposium XV*, 2003.

[35] http://www.intel.com/technology/mooreslaw/index.htm. Moore's laws. Technical report, Intel, 2005.

[36] http://www.itrs.net/Links /2005ITRS /Home2005.htm. The international technology roadmap for semiconductors. Technical report, Semiconductor Industry Association, 2005.

[37] http://www.mathworks.com/helpdesk/help/toolbox/quadprog. MathWorks quadprog function. Technical report, MathWorks, 2007.

[38] G. Hwaiyu. *Semiconductor Manufacturing Handbook*. McGraw-Hill, 2005.

[39] A. Ingolfsson and E. Saches. Stability and sensitivity of an EWMA controller. *Journal of Quality Technology*, 25-4:271–287, 1993.

[40] K. Joreskog, J. Klovan, and R. Reyment. *Geological Factor Analysis*. Elsevier, Amsterdam, 1976.

[41] A. Khuri. Applications of Dirac's delta function in statistics. *International Journal of Mathematical Education in Science and Technology*, 35(2):185–195, 2004.

[42] T. Kim and G. May. Intelligent control of via formation by photosensitive BCB for MCM-L/D applications. *IEEE Transactions on Semiconductor Manufacturing*, 12(4):503–515, 1999.

[43] S. Krishnan, G. W. Barton, and J. D. Perkins. Robust parameter estimation in online optimization 1: Methodology and simulated case study. *Computers and Chemical Engineering*, 16(6):545–562, 1992.

[44] G. Larabee. Flexible intelligent manufacture of microelectronic devices. *Manufacturing Review*, 7-3:212–226, 1994.

[45] M. Le, T. Smith, D. Boning, and H. Sawin. Run-to-run model based process control on a dual coil transformer coupled plasma etcher. In *191st Meeting of the Electrochemical Society*, 1997.

[46] S. Leang and C. Spanos. Statistically based feedback control of photoresist application. In *Proceedings of the IEEE/SEMI ASMC*, 1991.

[47] S. R. Leang, S. Y. Ma, J. Thomson, B. J. Bombay, and C. J. Spanos. A control system for photolithographic sequences. *IEEE Transactions on Semiconductor Manufacturing*, 9(2):191:207, 1996.

[48] S. Limanond, J. Si, and K. Tsakalis. Monitoring and control of semiconductor manufacturing processes. *IEEE Control Systems Magazine*, 18 (6):46–58, 1998.

[49] R. Mahadevan and F. J. Doyle. Efficient optimization approaches to nonlinear model predictive control. *International Journal of Robust and Nonlinear Control*, 13(3-4):309–329, 2003.

[50] T. E. Marlin and A. N. Hrymak. Real-time operations optimization of continuous processes. In *Proceedings of the Fifth Conference on Chemical Process Control (CPC V) Assessment and New Direction for Research*, 1997 93 156-164.

[51] H. Marten and T. Naes. *Multivariate Calibration.* Wiley, New York, 1989.

[52] G. May and S. Sze. *Fundamentals of Semiconductor Fabrication.* Wiley, 2003.

[53] G. S. May and C. J. Spanos. *Fundamentals of Semiconductor Manufacturing and Process Control.* IEEE: Hoboken, N.J. Wiley-Interscience, 2006.

[54] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis, 4th ed.* Wiley-Interscience, 2006.

[55] J. Moyne. Making the move to fab-wide APC. *Solid State Technology*, 47 (9):47, 2004.

[56] J. Moyne, E. Del Castillo, and A. M. Hurwitz. *Run-to-run Control in Semiconductor Manufacturing.* Boca Raton: CRC Press, 2001.

[57] J. Moyne, C. El Chemali, J. Kim, T. Parikh, J. Chapple-Sokol, J. Colt, R. Nadeau, and P. Smith. Gradient and radial uniformity of a CMP process utilizing a pre and post measurement strategy. In *Proceeding of CMP-MIC*, 2000.

[58] J. R. Moyne, N. Chaudhry, and R. Telfeyan. Adaptive extensions to a multibranch run-to-run controller for plasma etching. *Journal of Vacuum Science Technology*, 13 (3):1786–1791, 1995.

[59] J. A. Mullins, W. J. Campbell, and A. D. Stock. An evaluation of model predictive control in run to run processing in semiconductor manufacturing. In *Process, Equipment, and Materials Control in Integrated Circuit Manufacturing III SPIE, 182-189*, 1997.

[60] K. R. Muske and J. B. Rawlings. Model predictive control with linear models. *AIChE Journal*, 39(2):262–287, 1993.

[61] P. R. C. Nelson, P. A. Taylor, and J. F. MacGregor. Missing data methods in PCA and PLS: Score calculations with incomplete obesrvations. *Chemometrics and Intelligent Laboratory Systems*, 35:45–65, 1996.

[62] Y. Nishi and R. Doering. *Handbook of Semiconductor Manufacturing Technology*. Marcel Dekker, 2000.

[63] A. Papoulis. *Probability, Random Variables, and Stochastic Processes, Third Edition*. McGraw Hill, 1991.

[64] E. N. Pistikopoulos, V. Dua, N. A. Bozinis, A. Bemporad, and M. Morari. Online optimization via off-line parametric optimization tools. *Computers and Chemical Engineering*, 26(2):175–185, 2002.

[65] J. D. Plummer, M. D. Deal, and P. B. Griffin. *Silicon VLSI Technology*. Prentice Hall, 2000.

[66] S. J. Qin. Recursive PLS algorithms for adaptive data modeling. *Computers and Chemical Engineering*, 22(4-5):503–514, 1998.

[67] S. J. Qin, G. Cherry, R. Good, J. Wang, and C. A. Harrison. Control and monitoring of semiconductor manufacturing processes: Challenges and opportunities. In *Keynote at IFAC DYCOPS-7*, 2004.

[68] S. J. Qin, G. Cherry, R. Good, J. Wang, and C. A. Harrison. Semiconductor manufacturing process control and monitoring: A fab-wide framework. *Journal of Process Control*, 16 (3):179–191, 2006.

[69] S. J. Qin and R. Dunia. Determining the number of principle components for best reconstruction. *Journal of Process Control*, 10(2-3):245–250, 2000.

[70] S. J. Qin, G. Sheid, and T. Riley. Adaptive run-to-run control and modeling for a rapid thermal processor. *Journal of Vacuum Science and Technology B*, 21-1:301–310, 2003.

[71] S. J. Qin and T. Sonderman. From chemical process control to semiconductor manufacturing control. In *Keynote at the AEC/APC Symposium XIV*, 2002.

[72] A. E. Raftery, D. Madigan, and J. A. Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997.

[73] M. Ananda Rao and J. Srinivas. *Neural Networks: Algorithms and Applications*. Pangbourne, England: Alpha Science International, 2003.

[74] S. Ruegsegger, A. Wagner, J. S. Freudenberg, and D. S. Grimard. Feedforward control for reduced run-to-run variation in microelectronics manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 12-4, 1999.

[75] D. Ruppen, D. Bonvin, and D. W. T. Pippin. Implementation of adaptive optimal operation for a semi-batch reaction system. *Computers and Chemical Engineering*, 22(1/2):185–199, 1997.

[76] S. A. Russell, P. Kesavan, J. H. Lee, and B. A. Ogunnaike. Recursive data-based prediction and control of batch product quality. *AIChE Journal*, 44:2442–2458, 1998.

[77] E. Sachs, R. S. Guo, S. Ha, and A. Hu. Process control system for VLSI fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 4(2):134–144, 1991.

[78] E. Sachs, A. Hu, and A. Ingolfsson. Run by run process control: Combining SPC and feedback control. *IEEE Transactions on Semiconductor Manufacturing*, 8:26–43, 1995.

[79] C. Schoene and S. J. Qin. Electrical parameter control using latent variable models. In *AEC/APC Symposium XVII*, 2005.

[80] B. W. Silverman. *Density Estimation*. Chapman and Hall, 1986.

[81] T. Smith and D. Boning. A self tuning EWMA controller utilizing artificial neural network function approximation techniques. *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C*, 20-2:121–132, 1997.

[82] T. Smith and D. Boning. Enabling intermittent, delayed, and non-periodic data sampling with predictor corrector control. *Journal of Vacuum Science and Technology*, 2001.

[83] T. Smith, S. Fang, J. Stefani, G. Shinn, D. Boning, and S. Butler. Online patterned wafer thickness control of chemical mechanical polishing. *Journal of Vacuum Science Technology A*, 17(4):1384–1390, 1999.

[84] T. H. Smith, Boning D. S., J. Stefani, and Butler S. W. Run-by-run advanced process control of metal sputter deposition. *IEEE Transactions on Semiconductor Manufacturing*, 11(2):276–284, 1998.

[85] M.R Spiegel and J. Liu. *Mathematical Handbook of Formulas and Tables, 2nd Edition*. McGraw Hill, 1999.

[86] K. Stoddard, P. Crouch, M. Kozicki, and K. Tsakalis. Applications of feed-forward and adaptive feedback control to semiconductor device manufacturing. In *Proceedings of 1994 American Control Conference*, 1994.

[87] B. Streetman and S. Banerjee. *Solid State Electronic Devices*. Prentice Hall, 2005.

[88] M. Sullivan, S. W. Butler, J. Hirsch, and C. J. Wang. A control-to-target architechture for process control. *IEEE Transactions on Semiconductor Manufacturing*, 7(2):134–148, US Patent 5,402,367, 1994.

[89] S. M. Sze. *Physics of Semiconductor Devices 2nd Edition*. Wiley Interscience, 1981.

[90] R. Telfeyan, J. Moyne, N. Chaudhry, J. Pugmire, S. Shellman, D. Boning, W. Moyne, A. Hurwitz, and J. Taylor. A multi-level approach to the control of a chemical mechanical planarization process. *Journal of Vacuum Science and Technology A*, 14(3):1907–1913, 1996.

[91] S. T. Tseng and N. J. Hsu. Sample-size determination for achieving asymptotic stability of a double EWMA control scheme. *IEEE Transactions on Semiconductor Manufacturing*, 18(1):104–111, 2005.

[92] S. Vaidyanathan. *Stochastic Control of Sequential Manufacturing Processes*. PhD thesis, Carnegie Mellon University, 1991.

[93] J. Wang, Q. He, S. J. Qin, C. Bode, and M. Purdy. Recursive least squares estimation for run-to-run control of shallow trench isolation. *IEEE Transactions on Semiconductor Manufacturing*, 18-2:309–319, 2005.

[94] Z. Wang, S. Kwan, T. Pearsall, Booth J., Beard B., and S. Johnson. Real-time noninvasive temperature control of wafer processing based on diffusive reflectance spectroscopy. *Journal of Vacuum Science Technology B*, 15(1):116–121, 1997.

[95] S. Weisberg. *Applied Linear Regression.* New York, Wiley, 1980.

[96] S. Wold. Cross validatory estimation of the number of components in factor and principle component analysis. *Technometrics*, 20:397–406, 1978.

[97] S. Wold, C. Albano, W. J. Dunn III, K. Esbensen, S. Hellberg, E. Johansson, and M. Sjostrom. *Food Research and Data Analysis.* Applied Science Publishers, London and New York, 1983.

[98] S. Wold, K. Esbensen, and P. Geladi. Principle component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52, 1987.

[99] K. Wong, D. Boning, and H. H. Sawin. Endpoint detection for polysilicon plasma etch via optical emission interferometry. *Journal of Vacuum Science and Technology A*, 15(3):2531–2536, 1997.

[100] M. Yelverton, K. Tsakalis, and K. Stoddard. Factory-wide run-to-run process control. *Solid State Techonology*, 42(12):45, 1999.

[101] W. S. Yip and T. E. Marlin. Multiple data sets for model updating in real-time operations optimization. *Computers and Chemical Engineering*, 26(10):1345–1362, 2002.

[102] W. S. Yip and T. E. Marlin. The effect of model fidelity on real-time optimization performance. *Computers and Chemical Engineering*, 28(1-2):267–280, 2004.

# Index

# Vita

Clare Schoene was born in Santa Monica, CA on April 11, 1979 along with her twin sister Elizabeth Schoene. She attended Harvey Mudd College in 1997 and received her Bachelors of Science in Engineering degree in May 2001. She worked as a payload systems engineer for Boeing Satellite Systems for two years until she enrolled as a graduate student in chemical engineering at the University of Texas in August 2003.

Permanent address: 303 W 35th St Apt 204
Austin, Texas 78705

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.